GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:22:42 ; Search time 41.23 Seconds
                                  (without alignments)
                                  4212.634 Million cell updates/sec

Title: US-09-767-215-2
Perfect score: 5149
Sequence: 1 MGELCRRDSALTALDEETLW.........VRQAIADEQKKVVWTEQSPR 1004

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_19:*
        1: sp_archea:*
        2: sp_bacteria:*
        3: sp_fungi:*
        4: sp_human:*
        5: sp_invertebrate:*
        6: sp_mammal:*
        7: sp_mhc:*
        8: sp_organelle:*
        9: sp_phage:*
        10: sp_plant:*
        11: sp_rodent:*
        12: sp_virus:*
        13: sp_vertebrate:*
        14: sp_unclassified:*
        15: sp_rvirus:*
        16: sp_bacteriap:*
        17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5149	100.0	1004	4	Q9BXL6	Q9bxl6 homo sapien
2	2484.5	48.3	662	11	Q99KF0	Q99kf0 mus musculu
3	1910	37.1	434	4	Q9BVB5	Q9bvb5 homo sapien
4	1239.5	24.1	1147	4	Q9BXL7	Q9bxl7 homo sapien
5	887.5	17.2	1032	4	Q9BWT7	Q9bwt7 homo sapien
6	512.5	10.0	536	11	Q9EPY0	Q9epy0 rattus norv
7	493	9.6	536	4	Q9H257	Q9h257 homo sapien
8	396.5	7.7	366	4	Q9H854	Q9h854 homo sapien
9	351.5	6.8	1695	5	Q9BKL2	Q9bkl2 hydra atten
10	315	6.1	332	4	Q9UGR6	Q9ugr6 homo sapien
11	307.5	6.0	331	4	Q9UGR5	Q9ugr5 homo sapien
12	290.5	5.6	1769	6	Q97758	Q97758 canis famil
13	290	5.6	1313	4	O75033	O75033 homo sapien
14	282	5.5	974	5	Q960N4	Q960n4 drosophila
15	282	5.5	1367	5	Q94880	Q94880 drosophila
16	282	5.5	1445	5	Q9VHK4	Q9vhk4 drosophila
17	262	5.1	2101	4	Q14981	Q14981 homo sapien
18	259	5.0	813	11	P70625	P70625 rattus norv
19	258	5.0	1916	5	Q9VKG8	Q9vkg8 drosophila
20	254	4.9	904	11	Q921G9	Q921g9 mus musculu
21	253	4.9	1257	5	Q9NF04	Q9nf04 caenorhabdi
22	252	4.8	2115	4	Q14980	Q14980 homo sapien
23	248.5	4.8	2442	4	O14812	O14812 homo sapien
24	246	4.8	2611	11	Q91ZU8	Q91zu8 mus musculu
25	245	4.7	746	5	Q25561	Q25561 naegleria f
26	244.5	4.7	2442	4	Q9H450	Q9h450 homo sapien
27	243.5	4.7	2168	11	Q61043	Q61043 mus musculu
28	242.5	4.7	3259	4	Q14789	Q14789 homo sapien
29	242	4.7	1163	13	Q9HV2	Q9hv2 gallus gall
30	241.5	4.7	2442	4	O60588	O60588 homo sapien
31	236	4.6	1200	11	Q921B9	Q921b9 mus musculu
32	235.5	4.6	1692	4	Q9P216	Q9p216 homo sapien
33	235.5	4.6	1889	4	Q9H430	Q9h430 homo sapien
34	235	4.6	612	4	Q96JV2	Q96jv2 homo sapien
35	235	4.6	2029	4	Q9C014	Q9c014 homo sapien
36	235	4.6	2069	4	Q9C013	Q9c013 homo sapien
37	235	4.6	2073	4	Q9C012	Q9c012 homo sapien
38	233.5	4.5	1456	5	Q9V587	Q9v587 drosophila
39	230	4.5	1281	4	Q9Y4E3	Q9y4e3 homo sapien
40	229.5	4.5	2041	4	Q9HAT7	Q9hat7 homo sapien
41	229.5	4.5	2047	4	Q9UH61	Q9uh61 homo sapien
42	228	4.4	674	4	Q9UE73	Q9ue73 homo sapien
43	227.5	4.4	2473	11	Q9Q284	Q9q284 mus musculu
44	226.5	4.4	1270	4	Q96JN2	Q96jn2 homo sapien
45	223.5	4.3	2363	4	Q99968	Q99968 homo sapien

ALIGNMENTS

RESULT 1
Q9BXL6
ID   Q9BXL6   PRELIMINARY;   PRT;   1004 AA.
AC   Q9BXL6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CASPASE RECRUITMENT DOMAIN PROTEIN 14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21192234; PubMed=11278692;
RA   Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA   Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT   "CARD11 and CARD14 Are Novel Caspase Recruitment Domain
RT   (CARD)/Membrane-associated Guanylate Kinase (MAGUK) Family Members
RT   that Interact with BCL10 and Activate NF-kappaB.";
RL   J. Biol. Chem. 276:11877-11882(2001).
DR   EMBL; AF322642; AAG53403.1; -.
DR   InterPro; IPR000619; Guanylate_kin.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE 1004 AA; 113299 MW; 7469B8B56BE06073 CRC64;

Query Match   100.0%; Score 5149; DB 4; Length 1004;
Best Local Similarity   100.0%; Pred. No. 1.5e-295;
Matches 1004; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MGELCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHS 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MGELCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHS 60

```
DE   HYPOTHETICAL 75.4 KDA PROTEIN (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC004692; AAH04692.1; -.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   662 AA;  75422 MW;  AADFE37A7A04513D CRC64;

Query Match             48.3%;  Score 2484.5;  DB 11;  Length 662;
Best Local Similarity   75.1%;
Matches  505;  Conservative   63;  Mismatches   89;  Indels  89;  Gaps  ;
```

```
Qy     82 KNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGSLQEELNQ 141
          ||||||||||||||||||||||||||| ||||| ||||||||||||||||||| ||||| 
Db      1 KNGAIAFLESLKFHNPDVYTLVTGLQSDIDFSTFSGLMETSKLTECLAGAISSLQEELAQ  60

Qy    142 EKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRLKDEMLSL 201
          || ||||||||||||| ||||||| |||||||| |||||||||||| |||||||||||| 
Db     61 EKAQKEVLLRRCQQLKERLGLAEAHAEGLRQLEVDHSRMKREVSTHFHEVLKLKDEMLNL 120

Qy    202 SLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEELSRTASDQESGD 261
          ||||||| |||||| : |||||||||:|||||||: :|||||: :::  ::: :: :: :
Db    121 SLHYSNALREKELAAATRCHSLQEELYLVKQELQRASLVSSCERESRERSLKMASNLEPQG 180

Qy    262 EELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIHSLRERAVAAERQREQ 321
          ||||||||||||||::|||:|||||||||||||:||:||||:||||||||||||||:: :
Db    181 EELNRLKEENEKLRSMTFSLVEKDILEQSLDEARESKQELVDRIHSLRERAVAAERQQKQ 240

Qy    322 YWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYSARDSAQREISQSLVE 381
          |||||||||||| | : ||| :  :: ||:||| ||||||||||:|||:||:||||:|||
Db    241 YWEEKEQTLLQFRKTQVDCELYKEKMTMLQGQVAELQKERDQAYTARDRAQMEISQRLVE 300

Qy    382 KDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRD 441
          || |||:|||||:||||||||||:|||| ||  |||| :| : |  ||||||||| || :
Db    301 KDALRRRVFELTEQVCELRTQLRRLQAEAPGGPKQEAGARELCLRGKQRLVRMHAVCPPD 360

Qy    442 DSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSS 501
          ||||||:||||| | ||| :||||:: |||||||| ||||||||||||||| | :||   
Db    361 DSDCSLLSSTESRLWDLNSTSSREQMDSFRSSSPMPPSQQSLYKRVAEDFLEDPES---  417

Qy    502 CLEIPEGDPGALPGAKAG--DPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRR 559
           :  ||:| |||   ||    |  |:|::|:| ||||  ||||:||:| ||:| :|| ||
Db    418 -LSFPEVLEMRLQGATVDDTDTDLEFEMIDGADLSQTEDSLQG-SSRSLNVSESSVPVRR 475

Qy    560 RPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVD 619
          |||| ||||||:||||||||||||:||||||||||||||||||||| |||||||||:|||
Db    476 RPARKILSQVTVLAFQGDALLEQIGVIGGNLTGIFIHRVTPGSAADEMALRPGTQIMMVD 535

Qy    620 YEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLIQDLDAKVVTSGDSF 679
          | : :  : :: |::||| ||||||||:| | ||||:|:::| :||||||||||||||||
Db    536 YKPTKPSLRATLENTTLEQAVGLLRRVNGSCYLSVKINTEGYKNLIQDLDAKVVTSGDSF 595

Qy    680 YIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQQCGCWHAHRVNSYTMKDTAAHGTIPNYS 739
          ||||||| : | ||||:|||::|||||||||| | |||||:||:|||| ::  |||||||
Db    596 YIRVNLAMQRGGDGELQTHCNDILHVTDTMFQQGRSCWHAHHVNPYTMKDMEP-GTIPNYS 654

Qy    740 RAQQQLIALIQD 751
          :::  :|:|
Db    655 H------LLED 659
```

RESULT 3

```
Qy     61 PRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLME 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLME 120

Qy    121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180

Qy    181 KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240

Qy    241 SCELELQEELSRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQE 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 SCELELQEELSRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQE 300

Qy    301 LVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKE 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 LVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKE 360

Qy    361 RDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEART 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 RDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEART 420

Qy    421 REPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPS 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 REPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPS 480

Qy    481 QQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSL 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 QQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSL 540

Qy    541 QPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTP 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 QPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTP 600

Qy    601 GSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDG 660

Qy    661 YKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQQCGCWHAHR 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 YKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQQCGCWHAHR 720

Qy    721 VNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAK 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 VNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAK 780

Qy    781 ASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI 840

Qy    841 LSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTH 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 LSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTH 900

Qy    901 ALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDL 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDL 960

Qy    961 DRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQSPR 1004
          ||||||||||||||||||||||||||||||||||||||||||||
Db    961 DRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQSPR 1004
```

RESULT 2

```
Q99KF0
ID   Q99KF0            PRELIMINARY;      PRT;   662 AA.
AC   Q99KF0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
Q9BVB5
ID   Q9BVB5           PRELIMINARY;      PRT;    434 AA.
AC   Q9BVB5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 48.5 KDA PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CERVIX, CARCINOMA;
RA   Strausberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC001326; AAH01326.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE  434 AA;  48504 MW;  D351699781B6D6EB CRC64;

Query Match       37.1%;   Score 1910; DB 4;   Length 434;
Best Local Similarity  100.0%;   Pred. No. 6e-105;
Matches  381;   Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     238  MVSSCELELQEQSLRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGS  297
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1  MVSSCELELQEQSLRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGS   60

Qy     298  RQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCEL  357
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61  RQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCEL  120

Qy     358  QKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQE  417
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121  QKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQE  180

Qy     418  ARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPA  477
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181  ARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPA  240

Qy     478  PPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLE  537
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241  PPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLE  300

Qy     538  SSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHR  597
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301  SSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHR  360

Qy     598  VTPGSAADQMALRPGTQIVMV  618
            |||||||||||||||||||||
Db     361  VTPGSAADQMALRPGTQIVMV  381


RESULT  4
Q9BXL7
ID   Q9BXL7;           PRELIMINARY;      PRT;   1147 AA.
AC   Q9BXL7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CASPASE RECRUITMENT DOMAIN PROTEIN 11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21192234; PubMed=11278692;
RA   Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA   Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT   "CARD11 and CARD14 Are Novel Caspase Recruitment Domain
RT   (CARD)/Membrane-associated Guanylate Kinase (MAGUK) Family Members
RT   that Interact with BCL10 and Activate NF-kappaB.";
RL   J. Biol. Chem. 276:11877-11882(2001).
DR   EMBL; AF322641; AAG53402.1; -.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
SQ   SEQUENCE 1147 AA;  132641 MW;  913A4B015D2B36CC CRC64;

Query Match       24.1%;   Score 1239.5;  DB 4;   Length 1147;
Best Local Similarity  30.5%;   Pred. No. 7.9e-65;
Matches  354;   Conservative  204;  Mismatches  402;  Indels  201;  Gaps  33;

Qy      15  DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL   74
            :|:.|||. .|: .||::|| .|.:|||||||| .||||
Db      11  EEDALWENVECNRHMLSRYINPAKLTPYLRQCKVIDEQDEDEVLNAPMLPSKINRAGRLL   70

Qy      75  DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGS  134
            :.|: |.|:: :|||||:| :: : ||||  : |  .
Db      71  DILHTKQGRGYVVFLESLEFYYPELYKLVTGKEPTRRFSTIVVEEGHEGLTHFLMNEVIK  130

Qy     135  LQEELNQEKGQKEVLLRRCQQLQ-EHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLR  193
            ||::: |:| ::|::||| :| | |:| : :: ||. |.||| | ::||: |:::|||:|
Db     131  LQQQMAKDLQRCELLARLRQLEDEKKQMTLTRVE-LLTFQERYKMKEERDSYNDELVK    189

Qy     194  LKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEQSLRT  253
            :.|: .:|: .::..:.::.:: .| ::|.||||| ::|  ||
Db     190  VKDDNYNLAMRYAQLSEEKNMAVMRSRDQLEITDQLKHRLNK--MEEECKLE-RNQSLKL  246

Qy     254  ASDQES--GDEELNRLKEENEKLRSLITFSL---------------AEKDILEQSLDEAR  295
            ::.|:: :.::||.|:|||::|:||::.:|
Db     247  KNDIENRPKKEQVLELERENEMLKTKNQELQSIIQAGKRSLPDSDKAILDILEHDRKEAL  306

Qy     296  GSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVC  355
            .:.|||.||:::|::. :|::|.| ||.:|::|.:: :: .| ||||.| :::: :: :.:
Db     307  EDRQELVNRIYNLQEEARQAEELRDKYLEEKEDELKCSTLGKDCEMYKHRMNTVMLQLE   366

Qy     356  ELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQ----------------VCELRT  401
            ::::::|..:|:. :: . : .: : |::|::|:||:||
Db     367  EVERERDQAFHSRDEAQTQYSQCLIEKDKYRKQIRELEEKNDEMRIEMVRREACIVNLES  426

Qy     402  QLRQLQAE----------------PPGVLKQ-  428
            :||.||.: ||.::|
Db     427  KLRRLSKDSNNLDQSLPRNLPVTIISQDFGDASPRTNGQEADDSSTSEESPEDSKYFLPY  486

Qy     429  --QRLVRMHAI-CPRDDSDCSLVSSTESQLL-------------SDLSATS--SRELVDSFRSSSP  476
            | |:||.| | : : :|. |.|::|
Db     487  HPPQRRMNLKGIQLQRAKSPISLKRTSDFQAKGHEEESTDASPSSCGSLPITNSFTKMQP  546

Qy     477  -------------APPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKA-GDPHL  523
            ::|.::
Db     547  PRSRSSIMSITAEPPGNDSIVRRYKEDAPHR-----STVE-EDNDSGGFDALDLDDDSHE  600

Qy     524  DY------------------------ELLDTADLPQLESSLQPVS--PGRLDVSESGVLMRRRPARR  564
            : : ::: :.: ::| |..|
Db     601  RYSFGPSSIHSSSSSHQSEGLDAYDLEQVNLMFRKFSLERPFRPSVTSVGHVRGPGPS--  658

Qy     565  ILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASE  624
            :||.:.||.||. :|| .|: :|:.| ||.|.|: ||
Db     659  ---VQHTLNGDSLTSQLTLGGNARGSFVHSVKPGSLAEKAGLREGHQLLLLEGCIRG    714

Qy     625  PLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVN  684
            :|:.|:||.|:|||||||.||||||||:.|:|:.||| |.|:|||||||.|||||:||||
Db     715  ERQSVPLDTCTKEEAHWTIQRCSGPVTLHYKVNHEGYRKLVKDMEDGLITSGDSFYIRLN  774

Qy     685  LAMEGRAKG-ELQVHCNEVLHVTDTMFQGCGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQ  743
            |: .|.|| ::|.| |.:|:.:|::|.| ..: |.:.|.|||||:|||.|||||||||||
Db     775  LNISSQLDACTMSLKCDDVVHVRDTMYQDRHENPCARVDPFTDHDLDM-GTIPSYSRAQQ  833

Qy     744  QLIALIQDMTQQ-----CTVTRKP------KLV  771
            || :|: |: :| || | ||
Db     834  LLLVKLQRLMHRGSREEVDGTHHTLRALRNTLQPEEALSTSDPRVSPRLSRASFLFGQLL  893
```

```
Qy  341 QLYREKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELR  400
            :|: ::  :|  :|:::::|||||:: ::||   | : |::|:::||||:|: ||:
Db  375 DLYKHRMATVLAQLEEIEKERDQAIQSRDRIQLQYSQSLIEKDQYRKQVRGLEAERDELL  434

Qy  401 TQLRLQAEPPGVLKQEARTR-EPCPREKQRLVRMHAICPRDSDCSLVSST------     451
        |:| | ||||||:|:: |:           |     |:  |:|||:: |:|||::|
Db  435 TTLTSLEGTKALLEVQLQRAQGGTCLK-------ACASSHSLCSNLSSTWSLSEFPS    484

Qy  452 ------ESQLLSDLSATSSRELVDSFR-----SSSPAPPSQQSLYKRVAEDFGEEP-    496
              :: : :|:|| |:|  |||:|||  |   : ||:|| :||: || :  ::  |
Db  485 PLGGPEATGEAAVMGGPEPHNSEEATDSEKEINRLSILPFPPSAGSILRRQREEDPAPPK  544

Qy  497 WSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVL  556
        :||| |  ||     ::                      ||  :    | : |:::: :
Db  545 RSFSSMSDI----TGSV-------------TLKPWSPGLSSSSSSDSV              575

Qy  557 MRRRPARRILSQVTMLAFQGDALL-------EQISVIGGNLTGIFIHR            597
        ::|  ::   :|: :|  |||:: : |        :|:|||: : :||||||:| : :||
Db  576 WPLGKPEGLLARGCGLDFLNRSLAIRVSGRSPPGGPEPQDKGPDGLSFYGDRWSGAVVRR  635

Qy  598 VT-PGSA---ADQMALRPGTQIVMVDYEASEPLFKAVL--EDTTLEEAVGLLRRVDGF    649
        |  | |:   | :::    :|  : |  :   |  ::     :||| ||| :|||: :
Db  636 VLSGPGSARMEPREQRVEAAGLEGACLEAEAQQ---RTLLWNQGSTLPSL-----MDSK   686

Qy  650 CCLSVKVNTDGYKRLLQDLEAKVATSG-DSFYIRVNLAMEGRAKGE-LQVHCNEVLHVTD  707
          |  |:|| ||:|  |  |  || :   ||||||:|| : | || :  |:  ::|| ||
Db  687 ACQSFH----------EALEAWAKGPGAEPFYIRANLTLPERADPHALCVKAQEILRLVD  736

Qy  708 TMFQGCGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQOLIALIQDMTQQCTVTRKPSS--G  765
        : :   |||: | | |   | ||: |:|||| :||||:|      |:    ::| || | 
Db  737 SAYKRRQEWFCTRVDPLTLRD-LDRGTVPNYQRAQQLL-----EVQEKCL----PSSRHR  786

Qy  766 GPQKLVRIVSMD------KAKASPLRLSFPDRGQLDPSRMEGSSTCFWAESCLTLVPYTL  818
        ||::| |: :: |      :    |::   | ::  |: |         |    |||| |
Db  787 GPRSNLKKRALDQLRLVRPKPVGAPAGDSPDQLLEP--------C--AEPERSLRPYSL    836

Qy  819 VWPHRPARPRPVLLVPRAVGKILSEKLCLLQG---FKKCLAEYLSQEEYEAWSQRGDII   874
        | : |  ||||:||| :  |:| |:|| |||  :   |: |   |||||: | | |  :|
Db  837 VRPLLVSALRPVLLPECLAPRLIRNLLDLPSSRLDFQVCPAESLSGEELCPSSAPGAPK   896

Qy  875 QEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKK  934
        ::  :  |  | :|:| ::|: :   | |  : | |     ::  |||||| |:||:  :
Db  897 AQPATPGLGSRI-RAIQESVGKK--HCLLELGARGVRERVQNEIYPIVIHVEVTEKNVRE  953

Qy  935 LKKGLQRLGTSEEQLLEAARQEEGDLDRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQK  994
         ::||:|  |||:|||:| |: :  ||:  |  :  |||:|  :|||| |:|  :|:||:|
Db  954 VRGLLGRPGWRDSELLRQCRGSEQVLWGLPCSWVQVPAHEWGHAEELAKVVRGRIIQEQA 1013

Qy  995 KVVWTEQ                                                     1000
          :||:
Db 1014 RLVWVE                                                      1019
```

RESULT 6

```
Q9EPY0

ID   Q9EPY0
AC   Q9EPY0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CASPASE RECRUITMENT DOMAIN PROTEIN 9.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=20576268; PubMed=11053425;
RA   Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA   Poyet J.-L., Merriam S., Du M.Q., Dyer M.J.S., Robison K.E.,
RA   DiStefano P.S., Alnemri E.S.;
```

Q9EPY0   PRELIMINARY;   PRT;   536 AA.

```
Qy  772 RIVSMDKAK--------ASPL---RLSFD------RGQLDPSRMEGSSTCF          805
        ::|: || |          :|      ::         :|| |   | :
Db  894 QFVSRSENYKRMNSNERVRIISGSPLGSLARSSLDATKLLTEKQEELDPESELGKN--   950

Qy  806 WAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQG---FKKCLAEYLSQE  862
         : | : :|:|:|:|: |    |||| | ||| || ||::|  :      |::   :|:
Db  951 ----LSLIPYSLVRAFYCERRRPVLFTPTVLAKTLVQRLLNSGGAMEFTICKSDIVTRD 1005

Qy  863 EYEAWSQRGDII--QEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFP  920
        | | : ||  :|  ::  :       :| |  :: :|   | |:  :    : :   :
Db 1006 EFLRRQKTETIIYSREKNPNAFEC-IAPANIEAVAAKNKHCLLEAGIGCTRDLIKSNIYP 1064

Qy  921 IVIHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDLDRAPCLYSSLAPDGWSDLDG  980
        || : || || ::: : | :  |  |:|  : |::|::: |   |||::: :|:|  : :
Db 1065 IVLFIRVCEKNIKRFRKLLPRPET-EEEFLRVCRLIKEKELEALPCLYAIVEPDMWGSVEE 1123

Qy  981 LLSCVRQAIADEQKKVVWTEQ                                        1001
        || |:|: :   :::  | ::
Db 1124 LLRVVKDKIGEEQRKTIWVDE                                       1144
```

RESULT 5

```
Q9BWT7

ID   Q9BWT7
AC   Q9BWT7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CASPASE RECRUITMENT DOMAIN PROTEIN 10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21292987; PubMed=11259443;
RA   Wang L., Guo Y., Huang W.J., Ke X., Poyet J.L., Manji G.A.,
RA   Merriam S., Glucksmann M.A., DiStefano P.S., Alnemri E.S., Bertin J.;
RT   "CARD10 Is a Novel Caspase Recruitment Domain/Membrane-associated
RT   Guanylate Kinase Family Member That Interacts with BCL10 and Activates
RT   NF-kappa B.";
RL   J. Biol. Chem. 276:21405-21409(2001).
DR   EMBL; AY028896; AAK26165.1; -.
SQ   SEQUENCE 1032 AA; 115946 MW; 8377319AB82A0949 CRC64;
```

```
Query Match             17.2%;  Score  887.5;  DB  4;  Length 1032;
Best Local Similarity   29.2%;  Pred. No. 4e-44;
Matches  317;  Conservative  161;  Mismatches  419;  Indels  189;  Gaps  31;
```

```
Qy   15 DEETLWEMEMESHRHRIRVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRLTNSAMRAGHLL  74
         :|:|||::|||:|:|:|::|:|:|||||||| :: |:|||||| |: ||:|:: |:|||:
Db   23 EEDALWERIEGVRHRLARALRALNPAKLTPYLRQCRVIDEQDEEEVLSTYRFPCRVNRTGRLM  82

Qy   75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGS  134
         |:|: |||: | ||| :|:|::|:: ||:|| |:   |:: |: ::|:||  |||:||:
Db   83 DILRCRGRKKGYEAFLEALEFYYPEHFTLLTGQPEAQRCSMILDEEGPEGLTQFLMTEVRR  142

Qy  135 LQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRL  194
         |:| ::::|: :|| ||  :|:| | : :::|||||:||       |||:|  :||:|||
Db  143 LREARKSQLQREQQLQARGRVLEEERAGLEQRLRDQQQAQERCQRLREDWEAGSLELLRL  202

Qy  195 KDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEQSLRTA  254
         |||  : ::  :|| :: :  ::|  |::::::|||: | :    | ::||   :| | 
Db  203 KDENYMIAMRLAQLSEERNSAVLRSRDLQLAVDQLKLKVSR-LEEECAL----LRRA    254

Qy  255 SDQESGDEELNRLKEENE-------KLRS----LTFSLAE---                 283
           :   :||::| ::::         |         |:|
Db  255 RGPPPGAEEKEKEKEKEPDNVDLVSELRAENQOLTASLRELQEGLGQOEASRPGAPGSE  314

Qy  284 ---KDILEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMAC  340
            |||:|: :| |||:|||:| :|:| | ||||||:| |||||||: ||  :|:| : |
Db  315 RILLDILEHDWREAQDSRQELCQKIHAVGQELQWAEEELRDQYLQEMEDLRLKIHRTLQKDC 374
```

Left column

```
      "CARD9 is a novel caspase recruitment domain-containing protein that
      interacts with BCL10/CLAP and activates NF-kappa B.";
      J. Biol. Chem. 275:41082-41086(2000).
      EMBL; AF311288; AAG28791.1; -.
      InterPro; IPR001315; CARD.
      PROSITE; PS50209; CARD; 1.
      SEQUENCE 536 AA; 62631 MW; 6F33089CB7E6BAC9 CRC64;

Query Match         10.0%;  Score 512.5;  DB 11;  Length 536;
Best Local Similarity   28.8%;  Pred. No. 2.2e-22;
Matches  162;  Conservative   94;  Mismatches  215;  Indels   91;  Gaps  15;

Qy   15 DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL 74
            :: : :  :  ::  | ||:|||||| ||  ||||| |   : :|
Db    6 NDDECWSALESFRVKLISVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRRKVGVLL 65

Qy   75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPD------VDFSNFSGL-------ME 120
         :|: :|  |  ||||| |   |  | :|                    :|
Db   66 DILQRTGHKGYVAFLESLELYYPQLYRKVTGKEPARVFSMIIDASGESGLTQLLMTEVMK 125

Qy  121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
          :|:  | : :: |  :|  |:  || : |||  :||  :| |
Db  126 LQKKVQDLTALLSS-KDDFIKELRVKDSLLRKHQERVQ------------------RL 164

Qy  181 KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
          : | :   | | :|:||| :|  | :|:  :||    :| || ::|   |     |
Db  165 KEECELSSAELKRCKDENYDLAMRLAHLSEEKGAALMRNRDLQLEVDQLRHSLMKAE---D 222

Qy  241 SCELELQEQ-SLRTASDQESGDEELNRLKEENE---KLRSLTFSLAEK-------DI 286
         | :|  :: :|  | |     ::  |  :|            |:|| |
Db  223 DCKVERKHTLKLRHAMEQRPSQELLWDLQOQERDLLQARVQELEVSVQEGKLHRNSPYIQV 282

Qy  287 LEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREK 346
         ||: | : | :   |  |  |:|:||    :||   | || |  | |:|::   |||::|
Db  283 LEEDWRQALQEHQEQASTIFSLRKDLRQAEALRTRCMEEKEMFELQCLALRKDAKMYKDR 342

Qy  347 VNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQL 406
          :|:|  :  ||:|| : :::  |  || | :  :| ||:|||  | :|  ||  ||:|
Db  343 IEAILQQMEEVSIERDQAMTSREELHAQCAQSFQDKDKLRKQVRELDEKADELQLQLFQT 402

Qy  407 QAEPPGVLKQEARTREPCPREKQRIV--RMHAICPRDDSDCSLVSSTESQL-LSDLSAT 462
          : | | :|::: |   :   :|::|     : ||| :::||||| || |  |||
Db  403 ESR---LLAAEGRLKQ--QOLDMLILSSDLEDSSPRNSQELSLPQDLEEDAQLSDKGVL 456

Qy  463 SSRE--------LVDSFRSSSPAPPSQQSLYKRVAEDFG-------EEPWS 498
         | ||           | :: ||| |    | |: :|            ||
Db  457 ADRESPEQPFVVLNKKHLSQTHDTVPSSSEPPEKERRRLKESFENYRRRKALRKMQNSWR 516

Qy  499 FSSCLEIPEGDPGALPGAKAGD 520
            | | |: | | |
Db  517 QG------EGDHGNTTGSDNTD 532
```

```
RESULT 7
ID   Q9H257;       PRELIMINARY;   PRT;   536 AA.
AC   Q9H257;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CASPASE RECRUITMENT DOMAIN PROTEIN 9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20576268; PubMed=11053425;
RA   Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA   Poyet J.-L., Merriam S., Du M.Q., Dyer M.J.S., Robison K.E.,
RA   DiStefano P.S., Alnemri E.S.;
RT   "CARD9 is a novel caspase recruitment domain-containing protein that
```

Right column

```
RT   interacts with BCL10/CLAP and activates NF-kappa B.";
RL   J. Biol. Chem. 275:41082-41086(2000).
DR   EMBL; AF311287; AAG28790.1; -.
DR   InterPro; IPR001315; CARD.
DR   PROSITE; PS50209; CARD; 1.
SQ   SEQUENCE 536 AA; 62267 MW; 6EB18351315B83DE5 CRC64;

Query Match         9.6%;  Score 493;  DB 4;  Length 536;
Best Local Similarity   29.5%;  Pred. No. 3.2e-21;
Matches  155;  Conservative   92;  Mismatches  200;  Indels   78;  Gaps  14;

Qy   15 DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL 74
            :: :  ::  :  | ||||||| |||  ||||| |   : :|
Db    6 NDDECWNVLEGFRVTLTSVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRRKVGVLL 65

Qy   75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPD------VDFSNFSGL-------ME 120
         :|: :|  |  ||||| |   |  | :|                    :|
Db   66 DILQRTGHKGYVAFLESLELYYPQLYKKVTGKEPARVFSMIIDASGESGLTQLLMTEVMK 125

Qy  121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
          :|:  | : :: |  :|  |:  || : |||  :||  :| |
Db  126 LQKKVQDLTALLSS-KDDFIKELRVKDSLLRKHQERVQ------------------RL 164

Qy  181 KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
          : | :   | | :|:||| :|  | :|:  :||    :| || ::|   |     |
Db  165 KEECEAGSRELKRCKEENYDLAMRLAHQSEEKGAALMRNRDLQLEIDQLKHSLMKAE---D 222

Qy  241 SCELELQEQ-SLRTASDQESGDEELNRLKEE---NEKLRSLTFSLAEK-------DI 286
         | :|  :: :|  | |     ::  | :|            |:|| |
Db  223 DCKVERKHTLKLRHAMEQRPSQELLWELQQEKALLQARVQELEASVQEGKLDRSSPYIQV 282

Qy  287 LEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREK 346
         ||: | : | :   |  |  |:|:||    :||   | || |  | |:|::   |||::|
Db  283 LEEDWRQALRDHQEQANTIFSLRKDLRQGEARRLRCMEEKEMFELQCLALRKDSKMYKDR 342

Qy  347 VNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQL 406
          :|:|  :  ||:|| : :::  |  || | :  :| ||:|||  | :|  ||  ||:|
Db  343 IEAILLQMEEVAIERDQAIATREELHAQHARGLQEKDALRKQVRELGEKADELQLQVFQC 402

Qy  407 QAEPPGVLKQEARTREPCPREKQRLVRMHAI----CPRDDSDCSLVSSTESQLLSD----- 458
          : | | :|::: |   :   :|:           ||| :::||||| || |||
Db  403 EAQ---LLAVEGRLRR----QQLETLVLSSDLEDGSPRRSQELSLPQDLEDTQLSDKGCLA 456

Qy  459 --------LSATSSRELVDSFRS---SSPAPPSQQSLYKRVAEDF 492
                    : : :: | | ||     ||| |    | |: :|
Db  457 GGGSPKQPFAALHQEQVLRNPHDAGLSSGEPPEKER--RRLKESF 499
```

```
RESULT 8
ID   Q9H854;       PRELIMINARY;   PRT;   366 AA.
AC   Q9H854;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CDNA FLJ13939 FIS, CLONE Y79AA1000827.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira Y., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK024001; BAB14766.1; -.
```

us-09-767-215-2.rspt

```
DR   InterPro; IPR001315; CARD.
DR   PROSITE; PS50209; CARD; 1.
SQ   SEQUENCE   366 AA;  42980 MW;  FF995A0841C49216 CRC64;

Query Match             7.7%;  Score 396.5; DB 4; Length 366;
Best Local Similarity  30.8%;  Pred. No. 9.3e-16;
Matches  116;  Conservative  65;  Mismatches 146;  Indels  51;  Gaps  8;

Qy     15  DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRLTNSAMRAGHLL  74
               ::::|  : |  |||:||||||||   :  || : : |||
Db      6  NDDECWNVLEGFRVTLTSVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRKRKVGVLL  65

Qy     75  DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPD------VDFSNFSGL--------ME 120
               |: : |||: ||| :  || |||| |||||   ::| |
Db     66  DILQRTGHKGYVAFLESLELYYPQLYKKVTGKEPARVFSMIIDASGESGLTQLLMTEVMK 125

Qy    121  TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
               ||:| |||||||||||||||||||||||||||||||||||||||||||||||::: -RL 164
Db    126  LQKKVQDLTALLSS-KDDFIKELRVKDSLLRKHQERVQ---------------------RL 164

Qy    181  KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANVS 240
                |: : ||| :| : || |  :: :||  :|
Db    165  KEECEAGSRELKRCKENYDLAMRLAHQSEKGAALMRNRDLQLEIDQLKHSLMKAE--D 222

Qy    241  SCELELEQQ-SLRTASDQESGDEELNRLKEE---NEKLRSLTFSLAEK--------DI 286
               ::| |||  :||   | || || |||  |:   ||
Db    223  DCKVERKHTLKLRHAMEQRPSQELLWELQQEKALLQQARVQELEASVQEGKLDRSSPYIQV 282

Qy    287  LEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREK 346
               ||: :|| ||  : : |:||  |:||
Db    283  LEEDWRQALRDHQEQANTIFSLRKDLRQGEARRLRCMEEKEMFELQCLALRKDSKMYKDR 342

Qy    347  VNALQAQVCELQKERDQA 364
               : :| : : |||
Db    343  IEAILLQMEEVAIERDQS 360


DR   PROSITE; PS50002; SH3; 1.
SQ   SEQUENCE  1695 AA;  190878 MW;  AE0E7D09007316F1 CRC64;

Query Match             6.8%;  Score 351.5; DB 5; Length 1695;
Best Local Similarity  20.9%;  Pred. No. 3.3e-12;
Matches  240;  Conservative 170;  Mismatches 376;  Indels 363;  Gaps 49;

Qy     68  MRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTL-VTGLQPDVDF--------SNFSG 117
               :: :|  : |  :  ::                   : :
Db     67  LKVGDILISVNERNVDG----------RSHHDAVEALKAAGMEARMEIKRPSINPPKKNVDN 118

Qy    118  LMET----SKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQL 173
               |  :   |  ||| :                             ::
Db    119  LNDSKVNGSKETDSERGRSRKKPKEMEQESGKKH----------------HSSHQK 158

Qy    174  EADHSRMKREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQEL 233
               |||                                   ||
Db    159  EDDHDSEKPR---------------------------------HRSKSREKGEEGEKSSRS-- 186

Qy    234  QRANMVSSCELELQEQSLRTAS---------DQESGDE---------------ELNRLKEENEKLRSL 277
               :: :|||||:|: |::|||                  || :|
Db    187  HNKSLEDNADTDRSERSHRSRSKNREDSEGGDKSEKSHRSRSKNREDNAEGEKSHRSRSK 246

Qy    278  -------TFSLAEKDILEQSLDEARGSRQELVERIHSLRER-----------AVAAERQREQYWE 324
                                                          :
Db    247  NREDAETGEKSEKSHRSRSKNREDAETGEKSEKSHRSRSRNREDAETGEKSHRSRSKNRE 306

Qy    325  --EKEQTLLQFQKSKMACQLYR----------EKVNALQ--AQVCELQKERD-QAYSARDSA 371
               :|||||||||||||||:| :  || : :
Db    307  DPETEEKSLR-SKSKTREDLSKTTESEKSEKIPKSQSKASLGGMEKENSDRQKSKTREDA 365

Qy    372  QREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRL 431
               :: :| : |:|| |  :|  :
Db    366  GEEEKARIMSTQS-KPNIIESSDK-HRSRSKTRE---DTEGVEKPRSRSTQRKPNDMEKL 420

Qy    432  VRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRVAED 491
                                                                    |  ||:
Db    421  NSSRA--NEDGEKASSRSRSKSRSRDDLTRSKSKESL-HRSKSKEALNVDTI--KVVLN 474

Qy    492  FGEEPWSFSSCLEIPEGDPGA-LPGAKA------------GDPHLDYELLDTADL---- 533
               :: || ||                                    |
Db    475  RGDQGYGFSLGQQIFVKDLAKDSPAAKAKNLEKGDIVREINGTPLDNLKISECIELIRGA 534

Qy    534  ----------PQLESSLQPVSPGRLDVSE---------------SGV----------- 555
                  :
Db    535  SETLTLTIVKKPKLEQDVVPEKPKSLEVNDEKPKSKEPSSGGKKKQEVSEPSLKTKEQKMV 594

Qy    556  --------------------------------------------------------- 555
Db    595  HKEPDQKSSKITPPDRHSSPVPSNDNHNDTKDSWAEATPKKHFSSAAHPNQEVEDMNAKI 654

Qy    556  --LMRRRPARRILSQV---TMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALR 610
               : : : :           :|:  ||  : | ||
Db    655  ERLKSNRKLERRMSQLPNAKVISFHKTGSV-GIQVAGGNSVGIFVAAIRPDSAAAKEGLK 713

Qy    611  PGTQIVM---VDYEASEPLFKAVLEDTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQD 667
               | : :   |                |      |  ||  :
Db    714  PGDQIIMCNEIDF-------------ENITREEAVLILLALPDDVSLVVESKQSTFDQIKKE 762

Qy    668  LEAKVATSGDSFYIRVNLAMEGRAK-GELQVHCNEVLHVTDTMFQG-CGCWHAHRV-NS 723
               ||| : || :| ||||||:  |  :  |                ||
Db    763  L-------GDNFFIRVNFDHAEKANTNELTFRKGEIFNVRDTMYQGLIGYWQAQRVGKNA 815

Qy    724  YTMKDTAAHGTIPNYSRAQQLIALIQDMTQQCTVT---------RKPSSGGPQKLVRIVSM 776
                                                          :
Db    816  QMLE----RGVLPNKSRAEQ--LAIAQKMEERQTLTPSKSKLKRRNSIGGTLKKQKFVSQ 869

Qy    777  DKAKASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVL--LVP 834
               :                                                  |
Db    870  DRLD----ELSFTEGVQIPA------------YERVVLKVADF-----MRPVVLGPLAD 908

Qy    835  RAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQGRDIIQEGEVSGGRCWVTRHAVESL 894
```

Db　909 LARIKLLEE--------MPDRFELPSPSAYDHPGD--------NNSISLSEIRNV 947

Qy　895 MEKNTHALLDVQLDSVCTLHRMDIFPIVIHV-SVNEKMAKKLKKGLQRLGTSEEQLLEA- 952

Db　948 IAKNKHCLLDIVPEGIEMLMYAQLCPIVVMLNSPSRGAVKDMRQSLVK------EILSSP 1001

Qy　953 --------ARQEEGDLDRA------PCLYSSLAPDGWSDLDGLLSCVRQAIADEQ 993

Db　1002 TNVTNFTGDPLNAKQVKNLFNNANKLNEFYPHVFTSKINTHIAGGDVL----SRAFYEKL 1057

Qy　994 KKVVWTEQS 1002

Db　1058 KEVIFTQQS 1066

RESULT 10

Q9UGR6
ID　Q9UGR6;　　PRELIMINARY;　　PRT;　　332 AA.
AC　Q9UGR6;
DT　01-MAY-2000 (TrEMBLrel. 13, Created)
DT　01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT　01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE　DJ889J22B.1 (NOVEL PROTEIN (ISOFORM 2)) (FRAGMENT).
GN　DJ889J22B.1.
OS　Homo sapiens (Human).
OC　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX　NCBI_TaxID=9606;
RN　[1]
RP　SEQUENCE FROM N.A.
RA　Steward C.;
RL　Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR　EMBL; AL049851; CAB63075.1; -.
FT　NON_TER　　1　　1
SQ　SEQUENCE　332 AA;　37053 MW;　68234A527D009007 CRC64;

Query Match　　6.1%;　　Score 315;　　DB 4;　　Length 332;
Best Local Similarity　31.4%;　　Pred. No. 5.3e-11;
Matches 106;　Conservative　51;　Mismatches 145;　Indels　36;　Gaps　10;

Qy　677 DSFYIRVNLAMEGRAKGE-LQVHCNEVLHVTDTMFQGGCWHAHRVNSYTMKDTAAHGTI 735
Db　4 EPFYIRANLTLPERADPHALCVKAQEILRLVDSAYKRRQEWFCTRVDPLTLRD-LDRGTV 62

Qy　736 PNYSRAQQQLIALIQDMTQQCTVTRKPSS--GGPQKLVRIVSMD------KAKASPLRL 786
Db　63 PNYQRAQQLL-----EVQEKCL----PSSRHRGPRSNLKKRALDQLRLVRPKPVGAPAGD 113

Qy　787 SFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVSALRPVVLLPECLAPRLIRNLL 846
Db　114 SPDQLLLEP--------C--AEPERSLRPYSLVRPLLVSALRPVVLLPECLAPRLIRNLL 163

Qy　847 LLQG---FKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHAL 902
Db　164 DLPSSRLDFQVCPAESLSGEELCPSSAPGAPKAQPATPGLGSRI--RAIQESVGKQKHCL 221

Qy　903 LDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDLDR 962
Db　222 LELGARGVRELVQNEIYPIVIHVEVTEKNVREVRGLLGRPGWRDSELLRQCRGSEQVLWG 319

Qy　963 APCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTE 1000
Db　282 LPCSWVQVPAHEWGHAEELAKVVRGRIIQEQARLVWVE 319

RESULT 11

Q9UGR5
ID　Q9UGR5;　　PRELIMINARY;　　PRT;　　331 AA.
AC　Q9UGR5;
DT　01-MAY-2000 (TrEMBLrel. 13, Created)
DT　01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT　01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE　DJ889J22B.1 (NOVEL PROTEIN (ISOFORM 1)) (FRAGMENT).
GN　DJ889J22B.1.
OS　Homo sapiens (Human).
OC　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Homo.
OC　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX　NCBI_TaxID=9606;
RN　[1]
RP　SEQUENCE FROM N.A.
RA　Steward C.;
RL　Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR　EMBL; AL049851; CAB63076.1; -.
FT　NON_TER　　1　　1
SQ　SEQUENCE　331 AA;　36925 MW;　DFF0A6323E39F65E CRC64;

Query Match　　6.0%;　　Score 307.5;　　DB 4;　　Length 331;
Best Local Similarity　32.0%;　　Pred. No. 1.5e-10;
Matches 108;　Conservative　50;　Mismatches 143;　Indels　37;　Gaps　11;

Qy　677 DSFYIRVNLAMEGRAKGE-LQVHCNEVLHVTDTMFQGGCWHAHRVNSYTMKDTAAHGTI 735
Db　4 EPFYIRANLTLPERADPHALCVKAQEILRLVDSAYKRRQEWFCTRVDPLTLRD-LDRGTV 62

Qy　736 PNYSRAQQQLIALIQDMTQQCTVTRKPSS--GGPQKLVRIVSMD------KAKASPLRL 786
Db　63 PNYQRAQQLL-----EVQEKCL----PSSRHRGPRSNLKKRALDQLRLVRPKPVGAPAGD 113

Qy　787 SFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVSALRPVVLLPECLAPRLIRNLL 846
Db　114 SPDQLLLEP--------C--AEPERSLRPYSLVRPLLVSALRPVVLLPECLAPRLIRNLL 163

Qy　847 LLQG---FKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHAL 902
Db　164 DLPSSRLDFQVCPAESLSGEELCPSSAPGAPKAQPATPGLGSRI-RAIQESVGKK--HCL 220

Qy　903 LDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDLDR 962
Db　221 LELGARGVRELVQNEIYPIVIHVEVTEKNVREVRGLLGRPGWRDSELLRQCRGSEQVLWG 280

Qy　963 APCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTE 1000
Db　281 LPCSWVQVPAHEWGHAEELAKVVRGRIIQEQARLVWVE 318

RESULT 12

097758
ID　097758;　　PRELIMINARY;　　PRT;　　1769 AA.
AC　097758;
DT　01-MAY-1999 (TrEMBLrel. 10, Created)
DT　01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT　01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE　ZO-1 MDCK.
GN　ZO1-MDCK.
OS　Canis familiaris (Dog).
OC　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX　NCBI_TaxID=9615;
RN　[1]
RP　SEQUENCE FROM N.A.
RC　STRAIN=BREED: COCKER SPANIEL;
RX　MEDLINE=99196918; PubMed=10094817;
RA　Gonzalez-Mariscal L., Islas S., Contreras R.G., Garcia-Villegas M.R.,
RA　Betanzos A., Vega J., Diaz-Quinonez A., Martin-Orozco N.,
RA　Ortiz-Navarrete V., Cereijido M., Valdes J.;
RT　"Molecular characterization of the tight junction protein ZO-1 in MDCK
RT　cells.";
RL　Exp. Cell Res. 248:97-109(1999).
DR　EMBL; U55935; AAD11529.1; -.
DR　HSSP; P31016; 1BE9.
DR　InterPro; IPR000619; Guanylate_kin.
DR　InterPro; IPR001478; PDZ.
DR　InterPro; IPR001452; SH3.

```
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00625; Guanylate_kin; 1.
DR   Pfam; PF00595; PDZ; 3.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 3.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 3.
DR   PROSITE; PS50002; SH3; 1.
SQ   SEQUENCE   1769 AA;  197606 MW;  181E9F36CEBC96EF CRC64;

Query Match             5.6%;  Score 290.5;  DB 6;  Length 1769;
Best Local Similarity  23.6%;  Pred. No. 1.4e-08;
Matches  188;  Conservative 121;  Mismatches 288;  Indels 201;  Gaps 41;

Qy    247 QEQSLRTASDQESGDEE------------LNRLKEENEKLRSLTFSLAEKDILEQSLDEARG 296
              :!:||  ||: || :     |  |||:::|:| ||  |  |  ||:|||||||
Db    162 RERSLSPRSDRRSVASSQPPKPTKVTLVKSRKNEEYGLR-LASHIFVKEISQDSLAARDG 220

Qy    297 SRQE--LVERIH-SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQ 353
              |||   ||: | |:  || ||  | :| ||  :  | |:|||    ||  :|
Db    221 NIQEGDVVLKINGTVTENMSLTD-----------AKTLIERSKGKLKMVVQRDE----RAT 266

Qy    354 VCELQKERDQAYSARDSAQREIS--QSLVEKDSLRRQVFELTDQVCELRTQLRQLQOAEPP 411
               | |                   | |||  |    ||
Db    267 LLNVPDLSDSIHSANASERDDISEIQSLASDHSGRS--------------------HDRPP 307

Qy    412 GVLKQEARTREPCPREKQRLVRMHAICPRDDSDCCLVSSTESQLLSDLSATSSRELVDSF 471
            : | |     :||  |  || |::|    ||  |
Db    308 ----RHSRSRSPDQRSEPSDHSRHS--PQQPSSGSL-RSREERISKPGAV--------- 351

Qy    472 RSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTA 531
               ||   ||:   |:    | || |  |   ||  :|  |  ||  ||  :|  ||:
Db    352 --STPVKHADDHTHKTVEEVVVER--NEKQAPSLPEKP---VYAQVGQPDVDL------- 398

Qy    532 DLPQLESSLQPVSP--GRLDVS--ESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIG 587
              |||:||  ||: |   || :||     |   |   || :|::|  | |: ||
Db    399 --------PVSPSDGVLPNSTHEDGIL--RPSMKLVK----FRKGDSV--GLRLAG 438

Qy    588 GNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTL-------EEA 639
            | |:| | |: ||  |  | |  ||:||  |     |  |:||  |||  :   | ||
Db    439 GNDVGIFVAGVLEDSPAAKEGLEEGDQILRVN---NVDFTNIIREEAVLFLLDLPKGEEV 495

Qy    640 VGLLRRVDGFCCLSVKNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHC 699
            | || ||    |||      | |:||:     | |  ||  : :
Db    496 TILAQK--------KKDVVRRIVE-------SDVGDSFYIRTHFEYEKESPYGLSFNK 538

Qy    700 NEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTV 758
             |:| |||  |   |      ||:   |    |    || |  |: |: |  |  |: | :
Db    539 GEVFRVVDTLYNGKLGSWLAIRIGKNHKE--VERGIIPNKNRAEQ--LASVQ-------Y 587

Qy    759 TRKPSSGGPQKLVRIVSMDKAKASPLR-LSFDRGQLDPSRMEGSSTCFWAESCLTLVP-Y 816
            | | |:|| | |:      :|  |::  |  | | :|   ||  | | |    |  :|  |
Db    588 TLPKTAGG--------DRADFWRFRGLRSSKRNLRKSREDLS------AQPVQTKFPAY 632

Qy    817 TLVWPHRPARPRPVLLVPRAVGKILSEKICLLQGFKKCLAEVLSQEE--YE-AWSQRGD 872
              |  |  | | |:|  | ||  | |:  |: ||  | |   :|:      |||| |
Db    633 ERVVLREAGFLRPVTIF-GPIADVAREK----------LAREEPDIYQIAKSEPRD 677

Qy    873 IIQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSYCTLHRMDIFPIVIHVSVNEKMA 932
              |  | | |    | |       |  || | |      |  |  |   :|| :|
Db    678 AGTDQRSSG--IIRLHTIKQIIDQDKHALLDVTPNAVDRLNYAQWYPIVVFLNPDSKQG 734

Qy    933 KKL---------KKGLQRLGTSEEQLLEAARQEGDLDRAPCLYSSLAPDGWSDLDGLLS 983
             |:| |:||    |   |:|   :|:  |    |   | |   ||   ||  |:|:
Db    735 VKTMRMRLCPESRKSARKLYERSHKL----RKNNHHLFTTINLNSM-NDGW------YG 783

Qy    984 CVRQAIADEQKKVVWTEQ 1001
              :::||  :| :|: ||
```

```
Db    784 ALKEAIQQQNQLVWVSE 801

RESULT  13
075033
ID   O75033;                PRELIMINARY;       PRT;   1313 AA.
AC   O75033;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   KIAA0445 PROTEIN (DJ37C10.5) (KIAA0445).
DE   KIAA0445 OR DJ37C10.5.
GN   Homo sapiens (Human).
OS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=98116662; PubMed=9455484;
RA   Seki N., Ohira M., Nagase T., Ishikawa K., Miyajima N., Nakajima D.,
RA   Nomura N., Ohara O.;
RT   "Characterization of cDNA clones in size-fractionated cDNA libraries
RT   from human brain.";
RL   DNA Res. 4:345-349(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Heath P.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB007914; BAA23290.1; -.
DR   EMBL; AL049569; CAB96825.1; -.
SQ   SEQUENCE   1313 AA;  148254 MW;  00DC42FDE419EED1 CRC64;

Query Match             5.6%;  Score 290;  DB 4;  Length 1313;
Best Local Similarity  21.7%;  Pred. No. 9.9e-09;
Matches  246;  Conservative 165;  Mismatches 406;  Indels 318;  Gaps 51;

Qy    49 LCQLDEEVLHSPRLTNSAMRAGHL--LDLLKTRGKNGAI-------AFLESLKFHNP 97
            | :|||||  | :|:   || |   |  |: |:| ||:           ||| ||
Db     1 MLQAEKAEVAEA--LTKA-EAGRVLELSMTKLRAEEASLQDSLSKLSALNESLAQDKL 56

Qy    98 DVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGSLQEELNQPEKGQKEVLLRRCQQLQ 157
            : ||        |       |  | :| ||  || || ||:  ||  | ||||:||| |
Db    57 DLNRLVAQLEEK-----SALOGRQRAEQEATVAREEQERLEELRLEQEVAR----QGLE 108

Qy   158 EHLGLAETRAEGLHQLEADHSRMKREVSANHFEVLRLKDEMLSLSLHYSNALQEKELAAS 217
              :|| ||||| || | |  | : :| |  |  ||| | ||:|:|:  |  |:||
Db   109 -----------QLPTLRHERSQLEQLAQLSRQLSG--REQELEQA 156

Qy   218 RCRSLQEELYLL------KQELQRANMVSSCELELQEQSLRTASDQESGDEELNRLKEEN 271
             || |||: |         |||| |   |  | : |:|:| |  | ::| ||:|| :||
Db   157 R-REAQRQVEALERAAREKEALAKEHAGLAVQLVAAEREGRTLS------EEATRLRLEK 209

Qy   272 EKLRSLTF---------SLAEKDILEQSLD-------------EARGSRQELV 302
            |:| |  |         | ||:|  |:||               |||| ||:|
Db   210 EALEGSLFEVQRQLAQLEARREQLEAEGQALLLAKETLTGELAGLRQQIIATQEKASLDK 269

Qy   303 --------------ERIHSLRERAVAAE----------------RQREQYWEEKE---- 337
                          ||| |:|| | |                 |:|:|| |:||
Db   270 ELMAQKLVQAEREAQASLREQRAAHEEDLQRLQREKEAAWRELEAERAQLSQLQREQEE 329

Qy   338 MACQLYREKVNALQAQVCELQKERD------------QAYSARDSAQREISQSLV----- 380
            :| ||  :|  || | | | |:|:               || || :||:||:|:
Db   330 LLARLEAEK-EELSEEIAALQQERDEGLLLAESEKQQALSKESEKTALSEKLMGTRHSL 388

Qy   381 --------------VFELTDQVCELRTQ---------LRQLQAEPPGV 413
                          || |||:| |||          |||||:|  |:
Db   389 ATISLEMERQKRDAQSREQDRSTVNALTSELRDLRAQREEAAAAHAQEVRRLQEQARDL 448

Qy   414 LKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDS--- 470
            |:||||  |||  || :  |  || |   | |::  :: ::|||  | | :| |: |
Db   449 GKQ----RDSCLREAEEL-----------------------------RTQLRLLEDARDGLRRELLEAQRK 486
```

```
Qy  471  FRSSSPAPPSQQSLYKRVAEDFGE-----------EPWSFSSCLEIPEGDPGALPGAKAGD  520
              !! !  : :!!  !                 !!! ! !:::::::  !!
Db  487  LRESQEGREVQROEAGELRRSLGEGAKEREALRRSNEELRSAVKKAESERISLKLANE-D   545

Qy  521  PHLDYELLDTA------DLPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAF   574
              !!  ! :!    :  !! !  !  !! ::!: :!             !! :!
Db  546  KEQKLALLEEARTAVGKEAGELRTGLQEVERSRLEA-------RRELQELRRQMKMLDS   597

Qy  575  QGDALLEQISVIGGNL----------TGIFIHRVTPGSAADQMALRPGTQIVMVDYE     621
          !:! !! : !: !               !! ! !    :  : !  :   !:!!
Db  598  ENTRLGRELAELQGRLALGERAEKESRRETLGLRQRLLKGEASLEV-MRQELQVAQRKLQ  656

Qy  622  ASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYI  681
              !  ! !  :   !!      !!!         !! !    !  !: !!  :!:!!
Db  657  EQEGEFR------TRERRLLGSLEEARG---TEKQQLDHARGLELKLEAARAEAAE-LGL   706

Qy  682  RVNLAMEGRAKG--------ELQVHCNEVL-HVTDTMFQGCGCWHA------HRVNSYTM   726
          !::  !!!!!         !  !  !:       !:    !
Db  707  RLS-AAEGRAQQLEAELARVEVQRRAAEAQLGGLRSALRRGLGLGRAPSPAPRPVPGSPA  765

Qy  727  KDTAAHGTIPNYSRAQQOLIALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAKASPLRL  786
          :  :    !!! !                         !!! !!!
Db  766  RDAPAEGSGE-------GLNSPSTLECSPGSQOPPSPGPA---------TSPASP---    803

Qy  787  SFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLC  846
              !:! : !                         !:  !   !:
Db  804  ------DLDPEAVRGALREFLQEL-----------RSAQRERDEL--RTQTSALNRQLA   843

Qy  847  LLQGFKKCL--AEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHALL   903
           :  !          !!! !:!:        !! ! :!!:!      !:  !!:!!!
Db  844  EMEAERDSATSRARQLQKAVAESEEARRSV--DGRLSGVQAEL--ALQEESVRRSERER   898

Qy  904  DVQLDSVCTLHR-MDIFPIVIHVSVNEKMAK-----KLKKGLQRLGTSEEQLLEAA--   953
          !!!:!:!:!!!  !!!:!!!!!!!!!!!!       !::! !:!   !:
Db  899  RATLDQVATLERSLQATESELRAS-QEKISKMKANETKLEGDKRRL---KEVLDASESR  953

Qy  954  RQEEGDLDRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTE            1000
           !! :   !    !!! !!:!    :!  !!:!!!!!!!!!!!!!!!
Db  954  GLSDREAQAQALQDRVDSLQRSRL-----GLSDREAQAQALQDRVDSLQRQVADSE    1000
```

```
RESULT 14
Q960N4
ID   Q960N4;
AC   Q960N4:
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LD43161P.
GN   PYD.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E, George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051965; AAK93389.1; -.
SQ   SEQUENCE  974 AA;  107338 MW;  5FA3A7D7D0922C2E CRC64;
```

```
Query Match         5.5%;          Score 282;  DB 5;  Length 974;
Best Local Similarity  22.7%;       Pred. No. 2e-08;
  Matches  155;  Conservative  111;  Mismatches  276;  Indels  142;  Gaps  30;
```

```
Qy  373  REISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLV  432
          !!!   :::   :: !!::  !!! !!   : :!!       !!  :! !!
Db  365  KEISSK--AREQLNANGYSL--QEGDIITRIHNTNCGDTMSLKEAKKIIDGC-KERLNLV   419

Qy  433  RMHAICPRDD-SDCSLVSSTESQLLSDLSAT----------SSRELVDSFRSSSPAPPSQ  481
          !!       !                            !!
Db  420  VLRDITNQTAVSQLNLNNSASHQASGNIYATHQPQVSGCSSSNNNLEDPYLPGG-ASYSS  478

Qy  482  QSLYKVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQ   541
          !:!!       !  ::  : !                   !: : !
Db  479  QNLYVQPPTRTSNGFNINGNLN--DEKSNLTPRGRSRGP----IMDGVSLQQLD---R   528

Qy  542  PVSPGR-------LDV-------SESGVLM------------RRRPARRILSQVTMLAFQGDALL  580
          !!: !        !:!                         !!! !!!!!!
Db  529  PVTPTRGRSAAIDEPPRPPPRGSSGGAAQEDFYSSRRQLYEERQSAEPRFISFQKEGSV   588

Qy  581  EQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAV  640
           !! !!!!  !! !        :   !!  !  !!      :
Db  589  -GIRLTGGNEAGIFVTAVQPGSPASLQGLMPGDKILKVN------DMDMNGVTREEAV   639

Qy  641  GLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATS-GDSFYIRVNLAMEGRAKGELQVH   698
          !!!!!!  !!!   :         ! !!!!!!!!!  !!! :!
Db  640  LFL-------LSLQDRIDLIVQYCKEEYDEVVTNQRGDSFHIKTHFHCDNPSKGEMAFK   691

Qy  699  CNEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQOLIALIQDMTQQCT  757
          !:!! !!  : !       :  !     !   !!  !!!  !!!!:!!!!!!!!!!!!!!
Db  692  AGDVFRVIDTLHNGVGSWQVLKIGRGHQE--MQRGVIPNKSRAEEL-------ATAQFN   742

Qy  758  VTRKPSSGGFQKLVRIVSMDKAKASPLRLSFDRGQLDPSR----MEGSSTCFWAESCL   811
          !!    !!! !   :         !!!!!!!!!!!!
Db  743  ATKK--------EMNANESRGNFFRRRSTHRRSKSLSRENWDDVVFSDSIS            786

Qy  812  TLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRG  871
          !!      !!  !!  !!    !:!          !   !!
Db  787  KFPAYERVVLRHPGFVRPVVLF-GPVSDLARERLA--KDFPDKFSTPLQDDDKSA----   838

Qy  872  DIIQEGEVSGGRCWVTRHA-VESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEK  930
          !:!:!    :!!!      !!!!:!!!!!!!!!!!!!!!!:!!!:!!!!!!!!!!!!!!
Db  839  -------ATSGKCRIVRLSNIRDVMDRGKHALLDITPNAVDRLNYAQFYPVVIFLKTDSK  891

Qy  931  -MAKKIKKGLQRLG-TSEEQLLEAARQEEGDLDRAPCLYSSLAPDGW----------LERV  978
           !!:!!!! !!:!   :!                                    !!!!
Db  892  HVIKQLRHGLPKAAHKSSKKLLEQCQK------LERV                         936

Qy  979  DGLLSCVRQAIADEQKKVVWTEQS    1002
          !:   !   :!   :!  ::
Db  937  ESWYRKLRDSIDLQQSGAVMSES    960
```

```
RESULT 15
Q94880
ID   Q94880          PRELIMINARY;      PRT;  1367 AA.
AC   Q94880;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TAMA.
GN   PYD OR TAMOU OR CG9763.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=96312452; PubMed=8698238;
RA   Takahisa M., Togashi S., Suzuki T., Kobayashi M., Murayama A.,
RA   Kondo K., Miyake T., Ueda R.;
RT   "The Drosophila tamou gene, a component of the activating pathway of
RT   extramacrochaetae expression, encodes a protein homologous to
RT   mammalian cell-cell junction-associated protein ZO-1.";
RL   Genes Dev. 10:1783-1795(1996).
```

Fri Jun 14 08:13:31 2002

us-09-767-215-2.rspt

```
DR   EMBL; D83477; BAA11923.1; -.
DR   HSSP; Q12923; 3PDZ.
DR   FlyBase; FBgn0003177; pyd.
DR   InterPro; IPR000619; Guanylate_kin.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00625; Guanylate_kin; 1.
DR   Pfam; PF00595; PDZ; 3.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 3.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 3.
SQ   SEQUENCE   1367 AA;  148347 MW;  571C4566C6B68BF8 CRC64;
```

```
Query Match               5.5%;     Score 282;  DB 5;   Length 1367;
Best Local Similarity    22.7%;     Pred. No. 3.1e-08;
Matches  155;  Conservative 111;  Mismatches 276;   Indels 142;   Gaps  30;

Qy   373  REISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLV 432
          ::!!     :   !! : : :: :  : : ! :  !  : :  :: :  !! :  !  !!
Db   194  KEISSK--AREQLNANGYSL--QEGDITTRIHNTNCGDTMSLKEAKKIIDGC-KERLNLV 248

Qy   433  RMHAICPRDD-SDCSLVSSTESQLLSDLSAT-------SSRELVDSFRSSSPAPPSQ 481
          : !!      !  !!   : !! :!    :  !  ! :  !  -
Db   249  VLRDITNQTAVSQLNLNNSASHQASGNIYATHQPQVSGCSSNNNLEDPYLPGG-ASYSS 307

Qy   482  QSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQ 541
          !! !  ! :  :       :  :    :  !!::  :: !!!:   -
Db   308  QNLYVQPPTRTSNGPNINGNGLN--DEKSNLTPRGRSRGP-----IMDGVSLQQLD---R 357

Qy   542  PVSPGR-----LDV------SESGVLM------RRRPARRILSQVTMLAFQGDALL 580
          !!::     :!                !     : ::!  !!:  ::!!:  !!  !!
Db   358  PVTPTRGRSAAIDEPPRPPPPRGSSGGAAQEDFYSSRRQLYEERQSAEPRFISFQKEGSV 417

Qy   581  EQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAV 640
          :: : !!!  !!! : !!!! !     :     !  : !: :: :  :     !!!!
Db   418  -GIRLTGGNEAGIFVTAVQPGSPASLQGLMPGDKILKVN------DMDNGVTREEAV 468

Qy   641  GLLRRVDGFCCLSVKVNTDGYKRLIQDLEAKVATS--GDSFYIRVNLAMEGRAKGELQVH 698
          !!::     !   : :  :!!:         !!::   ::!! !!!:
Db   469  LFL--------LSLQDRIDLIVQYCKEEYDEVVTNQRGDSFHIKTHFHCDNPSKGEMAFK 520

Qy   699  CNEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCT 757
          !! :!!!    !!  !! !!  !     !  : :  :       !! !!!!  -
Db   521  AGDVFRVIDTLHNGVGSWQVLKIGRGHQE-MQRGVIPNKSRAEEL------ATAQFN 571

Qy   758  VTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSR------MEGSSTCFWAESCL 811
          !::!                  : :  ! :  :          !!!  :::!!
Db   572  ATKK-------------EMNANESRGNFFRRRRSTHRRSKSLSRENWDDVVFSDSIS 615

Qy   812  TLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRG 871
          !!:::!  !!!! !  :  !!     :!!:!! :    :   !  :! : ! !   !
Db   616  KFPAYERVVLRHPGFVRPVVLF-GPVSDLARERLA-KDFPDKFSTPLQDDDKSA-----667

Qy   872  DIQEGEVSGGRCWVTRHA-VESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEK 930
          :!!:! : : :   !!!!:   : !! !!!!!:  !  : !!:::!::   !:!! : !  !
Db   668  ------ATSGKCRIVRLSNIRDVMDRGKHALLDITPNAVDRLNYAQFYPVVIFLKTDSK 720

Qy   931  -MAKKLKKGLQRLG-TSEEQLLEAARQEEGDLDRAPCLYSSLAPDGW---------SDL 978
          : !!! ! ! ! !!    :  !!  : :!  !   -
Db   721  HVIKQLRHGLPKAAHKSSKKLLEQCQK----LERV--------WSHIFSTQIALSDE 765

Qy   979  DGLLSCVRQAIADEQKKVVWTEQS 1002
          :  !!  ! : !! :!  !! : !!
Db   766  ESWYRKLRDSIDLQQSGAVWMSES 789
```

Search completed: June 13, 2002, 09:25:36
Job time: 174 sec

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      June 13, 2002, 09:22:57 ; Search time 15.16 Seconds
                                        (without alignments)
                                   2564.276 Million cell updates/sec

Title:       US-09-767-215-2
Perfect score: 5149
Sequence:    1 MGELCRRDSALTALDEETLW..........VRQAIADEQKKVVWTEQSPR 1004

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:    105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters:        105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  SwissProt_40:*

  Pred. No. is the number of results predicted by chance to have a
  score greater than or equal to the score of the result being printed,
  and is derived by analysis of the total score distribution.

                        SUMMARIES

            %
Result      Query
No.    Score  Match Length DB  ID            Description
-----  -----  ----- ------ --  --            -----------
  1    5149  100.0  1004   1  CARE_HUMAN    Q9bx16 homo sapien
  2    3869.5  75.2   999   1  CARE_MOUSE    Q99kf0 mus musculu
  3    1239.5  24.1  1147   1  CARB_HUMAN    Q9bx17 homo sapien
  4     887.5  17.2  1032   1  CARA_HUMAN    Q9bwt7 homo sapien
  5     886.5  17.2  1021   1  CARA_MOUSE    P58660 mus musculu
  6     512.5  10.0   536   1  CAR9_RAT      Q9epy0 rattus norv
  7     493    9.6   536   1  CAR9_HUMAN    Q9h257 homo sapien
  8     289    5.6  1745   1  ZO1_MOUSE     P39447 mus musculu
  9     280    5.4  1167   1  ZO2_MOUSE     Q920u1 mus musculu
 10     270.5   5.3  1736   1  ZO1_HUMAN     Q07157 homo sapien
 11     256    5.0  1174   1  ZO2_CANFA     Q95168 canis famil
 12     254.5   4.9  1190   1  ZO2_HUMAN     Q9udy2 homo sapien
 13     249    4.8  4684   1  PLEI_HUMAN    Q15149 homo sapien
 14     246.5   4.8   905   1  ZO3_MOUSE     Q9qxy1 mus musculu
 15     240.5   4.7  1960   1  MYH9_HUMAN    P35579 homo sapien
 16     238.5   4.6   898   1  ZO3_CANFA     O62683 canis famil
 17     236.5   4.6  4473   1  PLEI_CRIGR    Q9j155 cricetulus
 18     236    4.6  3210   1  CENF_HUMAN    P49454 homo sapien
 19     235    4.6  1959   1  MYH9_CHICK    P14105 gallus gall
 20     234.5   4.6  4687   1  PLEI_RAT      P30427 rattus norv
 21     226.5   4.4  1976   1  MYHA_RAT      Q9jit0 rattus norv
 22     226    4.4  1972   1  MYHB_RABIT    P35748 oryctolagus
 23     225.5   4.4  2349   1  TPR_HUMAN     P12270 homo sapien
 24     225.5   4.4  2704   1  BPA1_HUMAN    Q03001 homo sapien
 25     223    4.3  1939   1  MYH1_HUMAN    P12882 homo sapien
 26     219    4.3  1972   1  MYHB_HUMAN    P35749 homo sapien
 27     218    4.2  2230   1  GOG4_HUMAN    Q13439 homo sapien
 28     217.5   4.2   933   1  ZO3_HUMAN     O95049 homo sapien
 29     217.5   4.2  1976   1  MYHA_HUMAN    P35580 homo sapien
 30     217    4.2  1972   1  MYHB_MOUSE    O08638 mus musculu
 31     215.5   4.2  1087   1  AKA9_RABIT    Q28628 oryctolagus
 32     214    4.2  1407   1  TRHY_RABIT    P37709 oryctolagus
 33     214    4.2  1941   1  MYH2_HUMAN    Q9ukx2 homo sapien
 34     212.5   4.1  1940   1  MYH3_RAT      P12847 rattus norv
 35     212    4.1  1957   1  YD86_SCHPO    Q10411 schizosacch
 36     212    4.1  1976   1  MYHA_BOVIN    Q27991 bos taurus
 37     211.5   4.1  1939   1  MYH4_HUMAN    Q9y623 homo sapien
 38     210.5   4.1   976   1  SCP1_HUMAN    Q15431 homo sapien
 39     209.5   4.1  1937   1  MYH8_HUMAN    P13535 homo sapien
 40     209    4.0  1938   1  MYHD_HUMAN    Q9ukx3 homo sapien
 41     207.5   4.0  1790   1  USO1_YEAST    P25386 saccharomyc
 42     207.5   4.0  2871   1  DESP_HUMAN    P15924 homo sapien
 43     207    4.0  1084   1  MYSS_RABIT    P02562 oryctolagus
 44     207    4.0  1325   1  G160_MOUSE    P55937 mus musculu
 45     206.5   4.0  1290   1  XCPC_XENLA    P50532 xenopus lae


                        ALIGNMENTS

RESULT    1
CARE_HUMAN
ID   CARE_HUMAN     STANDARD;      PRT;   1004 AA.
AC   Q9BXL6; Q9BVB5;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Caspase recruitment domain protein 14 (CARD-containing MAGUK protein
DE   2) (Carma 2).
GN   CARD14 OR CARMA2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21192234; PubMed=11278692;
RA   Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA   Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT   "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10
RT   phosphorylation and NF-kappaB activation.";
RL   FEBS Lett. 496:121-127(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   PubMed=11356195;
RA   Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RT   "CARD11 and CARD14 are novel caspase recruitment domain
RT   (CARD)/membrane-associated guanylate kinase (MAGUK) family members
RT   that interact with Bcl10 and activate NF-kappaB.";
RL   J. Biol. Chem. 276:11877-11882(2001).
RN   [3]
RP   ERRATUM.
RA   Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RL   FEBS Lett. 505:198-198(2001).
RN   [4]
RP   SEQUENCE OF 1-740 FROM N.A.
RC   TISSUE=Cervix, and Colon;
RA   Strausberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC       phosphorylation of Bcl10.
CC   -!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC       interaction.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Expressed in placenta. Also detected in HeLa
CC       S3 cells, but not in the other cancer cell lines tested.
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC   -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC       PROSITE, Pfam or SMART.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC     modified and this statement is not removed.  Usage by  and for commercial
CC     entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC     or send an email to license@isb-sib.ch).
CC     -----------------------------------------------------------------------
DR     EMBL; AF322642; AAG53403.1; -.
DR     EMBL; AY032927; AAK54453.1; -.
DR     EMBL; BC018142; AAH18142.1; -.
DR     EMBL; BC001326; AAH01326.1; ALT_INIT.
DR     InterPro; IPR000619; Guanylate_kin.
DR     InterPro; IPR001478; PDZ.
DR     Pfam; PF00595; PDZ; 1.
DR     SMART; SM00072; GuKc; 1.
DR     SMART; SM00228; PDZ; 1.
DR     PROSITE; PS50209; CARD; 1.
DR     PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR     PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR     PROSITE; PS50106; PDZ; 1.
KW     Coiled coil.
FT     DOMAIN       15    107       CARD.
FT     DOMAIN      128    409       COILED COIL (POTENTIAL).
FT     DOMAIN      568    658       PDZ.
FT     DOMAIN      858    990       GUANYLATE KINASE.
FT     CONFLICT    619    671       DYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTD
FT                                  GYKRLLQDLEAK -> SRARPLLSPGLLMGTVAAGGVTQAD
FT                                  FTSPRRCRSTLGWASALSWADVKRSAHL (IN REF. 2;
FT                                  AAH01326).
SQ     SEQUENCE   1004 AA;  113299 MW;  7469B8B56BE06073 CRC64;
```

```
Query Match             100.0%;  Score 5149;  DB 1;  Length 1004;
Best Local Similarity   100.0%;  Pred. No. 5.4e-250;
Matches 1004;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MGELCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MGELCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHS 60

Qy     61 PRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLME 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLME 120

Qy    121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180

Qy    181 KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240

Qy    241 SCELELQEQSLRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 SCELELQEQSLRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQE 300

Qy    301 LVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKE 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 LVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKE 360

Qy    361 RDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEART 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 RDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEART 420

Qy    421 REPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPS 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 REPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPS 480

Qy    481 QQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSL 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 QQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSL 540

Qy    541 QPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 QPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTP 600
```

```
Qy    601 GSAADQMALRPCTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GSAADQMALRPCTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDG 660

Qy    661 YKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQGCGCWHAHR 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 YKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQGCGCWHAHR 720

Qy    721 VNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAK 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 VNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAK 780

Qy    781 ASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI 840

Qy    841 LSEKICLLQGFKKCLAEYLSQEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTH 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 LSEKICLLQGFKKCLAEYLSQEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTH 900

Qy    901 ALLDVQLDSVCTLHRMDIFPIVHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDL 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ALLDVQLDSVCTLHRMDIFPIVHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDL 960

Qy    961 DRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQSPR 1004
          |||||||||||||||||||||||||||||||||||||||||||||
Db    961 DRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQSPR 1004
```

```
RESULT    2
CARE_MOUSE
ID   CARE_MOUSE     STANDARD;      PRT;    999 AA.
AC   Q99KF0;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Caspase recruitment domain protein 14 (Bcl10-interacting MAGUK protein
DE   2) (Bimp2).
GN   CARD14 OR BIMP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21391892; PubMed=11387339;
RA   McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA   Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA   Nunez G.;
RT   "Bimp1, a MAGUK family member linking protein kinase C activation to
RT   Bcl10-mediated NF-kappa B induction.";
RL   J. Biol. Chem. 276:30589-30597(2001).
RN   [2]
RP   SEQUENCE OF 82-743 FROM N.A.
RC   TISSUE=Breast;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC   -!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC       interaction (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC   -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC       PROSITE, Pfam or SMART.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   use by and for commercial
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb.ch/announce/
```

```
CC        or send an email to license@isb-sib.ch).
CC
DR   EMBL; AF363457; AAK60137.1; -.
DR   EMBL; BC004692; AAH04692.1; -.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50209; CARD; 1.
DR   PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Coiled coil.
FT   DOMAIN        15     107       CARD.
FT   DOMAIN       125     411       COILED COIL (POTENTIAL).
FT   DOMAIN       572     655       PDZ.
FT   DOMAIN       854     986       GUANYLATE KINASE.
FT   CONFLICT     736     743       QAQQQLLA -> HLLEDHRS (IN REF. 2).
SQ   SEQUENCE    999 AA;  113496 MW;  D18350DA12430255 CRC64;
```

```
Query Match             75.2%;  Score 3869.5;  DB 1;  Length 999;
Best Local Similarity   76.9%;  Pred. No. 4e-186;
Matches  774; Conservative  83; Mismatches 138; Indels  11; Gaps   7;

Qy      1  MGELCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHS   60
           | |||| ||||||| || ||||| :||:|| ||| ||||||||||||||| :|||||||||
Db      1  MAELCRMDSTLTALDEEMLWDMLESHRCRIVQSICPSRLTPYLRQAKVLGQLDEEEILHS   60

Qy     61  PRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLME  120
           ||||||||| ||||||||:| |||||||||||||||||||||||||||||||| ||||||
Db     61  SRFTNSAMRVGHLLDLLKARGKNGAIAFLESLKFHNPDVYTLVTGLQSDIDFSTFSGLME  120

Qy    121  TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM  180
           |||||||||||| ||||||| ||| |||||||||:|||:|||| || ||:|| ||||||
Db    121  TSKLTECLAGAISSLQEELAQEKAQKEVLLRRCQQLKERLGLAEAHAEGLRQLEVDHSRM  180

Qy    181  KREVSAHFHEVLRLKDEMLSLSHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS  240
           ||||| ||||||| ||||| ||||||||||||::|||| ||||| |::||:||||:||
Db    181  KREVSTHFHEVLKLKDEMLNLSLHYSNALREKELAATRCHSLQEELYVLVKQLQRASLVS  240

Qy    241  SCELELEQSLRTAGSDEQESGDEELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQE  300
           ||| |:|: ||: : :| |||||||||||||||||:||:|| ::|| ||||||:|| ||:
Db    241  SCERESRERSLKMASNLEPQGEELNRLKEENEKLRSMTFSLVEKDILEQSLDEARESKQE  300

Qy    301  LVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKE  360
           |:|||||||||||||||:||:||||||||||||| ||:|| :|::||:||||||||||||
Db    301  LVDRIHSLRERAVAAERQQQYWEEKEQTLLQFRKTQVDCELYKEKMTMLQGQVAELQKE  360

Qy    361  RDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLROLQAEPPGVLKQEART  420
           |||| |:|| :|||:||| ||||:|| ||||||:||||||||:||| |||| ||||||||
Db    361  RDQAYTARDRAQMEISQRLVEKDALRRRVFELTEQVCELRTQLRRLQAEAPGGPKQEAGA  420

Qy    421  REPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPS  480
           |:|| ||:|||||||| ||||||||||| ||:|||||||:|:||||||||||||:| ||
Db    421  RELCLRGKQRLVRMHAVCPPDDSDCSLLSSTESRLWWDLNSTSSREQMDSFRSSSPMPPS  480

Qy    481  QQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAG--DPHLDYELLDTADLPQLES  538
           |||||||||||| : |:|| ||:||||| || ||:|| |  : |: || :||||:| :||
Db    481  QQSLYKRVAEDFLEDPES----LSFPEVLEMRLQGATVDDTDTDLEFEMIDCADLSQTED  536

Qy    539  SLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRV  598
           ||| :||| |||||| ||:||||:||||||||||||||||||||||||||||||||||||
Db    537  SLQG-SSRSLNVSESSVPVRRRPARKILSQVTVLAFQGDALLEQIGVIGGNLTGIFIHRV  595

Qy    599  TPGSSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNT  658
           |||||||: |||||||||||||||:||||| ||||||||||||||||||:||||||||:|
Db    596  TPGSSAADEMALRPGTQIMMVDYDYKPTKPSLRATLENTTLEQAVGLLRRVNGSCYLSVKINT  655

Qy    659  DGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQGCGCWHA  718
           : | ||||| ::| |:|||||:||||:||:| :||::|| |||:|||||||||||||: ||
Db    656  EGYKNLIQDLDAKVVFSGDSFYIIRVNLAMQRGGDGELQTHCNDILHVTDTMFQGRSCWHA  715

Qy    719  HRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSSGGPQKLVRIVSMDK  778
```

```
Db    716  HHVNPYTMKDMDEP-GTIPNYSQAQQQLALIQDMTQRCTVPRKP-PGGPQKLVRIVSVDK  773
           | || |||||||||  ||||||:|||||| ||||||:|||:|||  ||||||||||||:||
Qy    779  AKASPLRLSFDRGQLDPSRMEGS-STCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAV  837
           || |||| |||| :| ||:||:   |||| | ||:|:|||:||||:|||||||:|||||
Db    774  AAVSPLTSSFDQSQWDSGKEEGGPSVCFWSESCFTLAPYTLVHPHRPARPRPVLFVPRLV  833
Qy    838  GKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEK  897
           |:||| ||||||| | ||:||| |||||| |||||||||| |:  ||| |||||||||:|
Db    834  GRIGKKLCLLQGFKQCSAEYLSQEYATWSQRGDIIQEGSIGDHHWITRHAVESLMNM  893
Qy    898  NTHALLDVQLDSVCTLHRMDIFPYIHVSVNEKMAKKLKGLQRLGTSEEQLLEAARQEE  957
           :||||||| ||||||||||| |:|:||||||:|||||||:||||||:|||||| |||||
Db    894  STHALLDVRLDSVRVLHRMDMFPIIIHVSVNEKTAKKLRKGLHRLGSSEEQFLEVARQEE  953
Qy    958  GDLDRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQSP 1003
           ||| |||||||:||||:|||| :||||||||||||:|||||||| ||
Db    954  GELDRVPCLYSSLAPDSWSDLDSLLSCVRLAIADEDQKKVVWTE-SP  998
```

```
RESULT   3
CARB_HUMAN
ID   CARB_HUMAN    STANDARD;    PRT;   1147 AA.
AC   Q9BXL7;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Caspase recruitment domain protein 11 (CARD-containing MAGUK protein
DE   3) (Carma 1).
GN   CARD11 OR CARMA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21192234; PubMed=11278692;
RA   Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA   Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT   "CARD11 and CARD14 are novel caspase recruitment domain
RT   (CARD)/membrane-associated guanylate kinase (MAGUK) family members
RT   that interact with Bcl10 and activate NF-kappaB.";
RL   J. Biol. Chem. 276:11877-11882(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   PubMed=11356195;
RA   Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RT   "Carma1, a CARD-containing binding partner of Bcl10, induces Bcl10
RT   phosphorylation and NF-kappaB activation.";
RL   FEBS Lett. 496:121-127(2001).
RN   [3]
RP   ERRATUM.
RA   Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RL   FEBS Lett. 505:198-198(2001).
CC   -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC       phosphorylation of Bcl10.
CC   -!- SUBUNIT: CARD11 and Bcl10 bind to each other by CARD-CARD
CC       interaction.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Detected in adult peripheral blood leukocytes,
CC       thymus, spleen and liver. Also found in promyelocytic leukemia HL-
CC       60 cells, chronic myelogenous leukemia K562 cells, Burkitt's
CC       lymphoma Raji cells and colorectal adenocarcinoma SW480 cells. Not
CC       detected in HeLa S3, Molt-4, A549 and G431 cells.
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC   -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC   -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC       PROSITE, Pfam or SMART.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; AF322641; AAG53402.1; -.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50209; CARD; 1.
DR   PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; FALSE_NEG.
DR   PROSITE; PS50106; PDZ; FALSE_NEG.
KW   Coiled coil.
FT   DOMAIN      11    103       CARD.
FT   DOMAIN     123    442       COILED COIL (POTENTIAL).
FT   DOMAIN     673    748       PDZ.
FT   DOMAIN     966   1133       GUANYLATE KINASE.
FT   CONFLICT   808    808       P -> L (IN REF. 2).
SQ   SEQUENCE   1147 AA;  132641 MW;  913A8015D2B36CC CRC64;

Query Match             24.1%;  Score 1239.5;  DB 1;  Length 1147;
Best Local Similarity   30.5%;  Pred. No. 9.5e-55;
Matches  354;  Conservative  204;  Mismatches  402;  Indels  201;  Gaps  33;

Qy         15 DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRLTNSAMRAGHLL 74
              :||: ||  :.| |||| |: |||||||||  : ||:||:|||||: | |||:|||| ||
Db         11 EEDALWENVECNRHMLSRYINPAKLTPYLRQCKVIDEQDEDEVLNAPMLPSKINRAGRLL 70

Qy         75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGS 134
              | | |:| .|.:.||||| | .|::| |||| :| . : | |:.. ::   | |:| :.|
Db         71 DILHTKGQRGYVVFLESLEFYYPELYKLVTGKEPTRRFSTIVVEEGHEGLTHFLMNEVIK 130

Qy        135 LQEELNLQEKGQKEVLLRRCQQLQ-EHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLR 193
              || :: .:.| |.:.|| |  ||: | |.|.|:| : |: ::.:  :.||:.  :.:|:||
Db        131 LQQQMAKDLQRCELLARLRQLEDEKKQMTLTRVE-LLITFQERYYKMKEERDSYNDELVK 189

Qy        194 LKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEQSLRT 253
              :|:: . ::|: :||::|| ::| ||:.|| |: ||:.:::..:|  | | :|.:|.|:|
Db        190 VKDDNYNLAMRYAQLSEEKNMAVMRSRDLQLEIDQLKHRLNK--MEEECKLE-RNQSLKL 246

Qy        254 ASDQES--GDEELNRLKEENEKLRSLTFSL------------------AEKDILEQSLDEAR 295
              .  ::.  |:|:|: ||.::| |::: . .                  |:||||:.  :.|
Db        247 KNDIENRPKKEQVLELERENEMLKTKNQELQSIIQAGKRSLPDSDKAILDILEHDRKEAL 306

Qy        296 GSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVC 355
              . |||| .|| :|:|.| .||:|.| | |||:: .|:|:||  ||:||:||||.||.::.
Db        307 EDRQELVNRIYNLQEEARQAEELRDKYLEEKEDLELKCSTLGKDCEMYKHRMNTVMLQLE 366

Qy        356 ELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQ----------VCELRT 401
              |::|||||| .:||:||.: ||.|:||| ::|:: |:                :| :
Db        367 EVERERDQAFHSRDEAQTQYSQCLIEKDKYRKQIRELEEKNDEMRIEMVRREACIVNLES 426

Qy        402 QLRQLQAE----------PPGVLKQ----------EARTREPCPREK------ 428
              :||:|  :          ||.|::.          ||:|  ||| |
Db        427 KLRRLSKDSNNLDQSLPRNLPVTIISQDFGDASPRTNGQEADDSSTSEESPEDSKYFLPY 486

Qy        429 --QRLVRMHAI-CPRDDSDCCSLVSSTESQLL------SDLSATS--SRELVDSFRSSSP 476
                ||:. :.:: |.:.::  .|:.| |:|.:       |:||:|:  ||||||||:||||
Db        487 HPPQRRMNLKGIQLQRAKSPISLKRTSDFQAKGHEEEGTDASPSSCGSLPITNSFTKMQP 546

Qy        477 -------APPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDGALPGAKA-GDPHL 523
                     |||.  : ::|: .:|.||||.|::| ||||||  ||.::| ||:|
Db        547 PRSRSSIMSITAEPPGNDSIVRYKEDAPHR----STVE-EDNDSGGFDALDLDDDSHE 600

Qy        524 DY----------ELLTADLPQLESSLQPVS---PGRLDVSESGVLMRRRPARR 564
              ||          ||| .: .|::.:|.     ||||||||||:||||||.:||
Db        601 RYSFGPSSIHSSSSSHQSEGLDAYDLEQVNLMFRKFSLERPFRPSVTSVGHVRGPGPS-- 658

Qy        565 ILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASE 624
              :|:|:|:|  :||:| |::.|:|||: |.|:|.|.|||:|::.||| |:|::::||    .
Db        659 ----VQHTTLNGDSLTSQLTLLGGNARGSFVHSVKPGSLAEKAGLREGHQLLLEGCIRG 714

Qy        625 PLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVN 684
              .: :|||:|  :|||..:::|:|. | :.::|||:|||:|::||| ::.|||||||||:|
Db        715 ERQSVPLDTCTKEEAHWTIQRCSGPVTLHYKVNHEGYRKLVKDMEDGLITSGDSFYIRLN 774

Qy        685 LAMEGRAKG-ELQVHCNEVLHVDTMFQGCGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQ 743
              |.: .: .  ::  .| ::|||.||.|.   |.|.||: :|:.|:. : -||||||:|||
Db        775 LNISSQLDACTMSLKCDDVVHVRDTMIQDRHEWPCARVDPFTDHDLDM-GTIPSYSRAQQ 833

Qy        744 QLIALIQDMTQQ-----------CTVTRKP----SSGPQ------------KLV 771
              .|: ::: : : :           .|.|:|    ||.:.              |:
Db        834 LLLVLKQRLMHRGSREEVDGTHTLRALRNTLQPEEALSTSDPRVSPRLSRASFLFGQLL 893

Qy        772 RIVSMDKAK----------ASPL--RLSFD---------RGQLDPSRMEGSSTCF 805
              :|||: :.:          ||.:  ||| :         ||| || ::||:|:
Db        894 QFVSRSENKYKRMNSNERVRIISGSPLGSLARSSLDATKLLTEKQEELDPESELGKN--- 950

Qy        806 WAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKICLLQG--FKKCLAEYLSQE 862
               |.| |: :||:||.    :.|||||| .|:| ::|.:::|| ||  ||:||||: | :
Db        951 -----LSLIPYSLVRAFYCERRRPVLFTPTVLAKTLVQRLLNSGGAMEFTICKSDIVTRD 1005

Qy        863 EYRAWSQRGDII--QEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFP 920
              |:   :::.  :  :.. .  : | : :: . ::::|| :  :  : . : |.:. : .|
Db       1006 EFLRRQKTETIIYSREKNPNAFEC-IAPANIEAVAAKNKHCLLEAGIGCTRDLIKSNIYP 1064

Qy        921 IVIHVSVNEKMAKKIKKGLQRLGTSEEQLLEAARQEEGLDRAPCLYSSLAPDGWSDLDG 980
              || .: |.|::.:.::|  ::|.  :||:|  .| :.|| |.|||:::: ||.|  : |
Db       1065 IVLFIRVCEKNIKRFRKLLPRPET-EEEFLRVCRLKEKELEALPCLYATVEPDMWGSVEE 1123

Qy        981 LLSCVRQAIADEQKKVVWTEQ 1001
              ||  |::|| :|| | ||::|
Db       1124 LLRVVKDKIGEEQRKTIWVDE 1144
```

```
RESULT 4
CARA_HUMAN

ID   CARA_HUMAN     STANDARD;      PRT;  1032 AA.
AC   Q9BWT7; Q9UGR5; Q9UGR6; Q9Y3H0;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Caspase recruitment domain protein 10 (CARD-containing MAGUK protein
DE   3) (Carma 3).
GN   CARD10 OR CARMA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21292987; PubMed=11259443;
RA   Wang L., Guo Y., Huang W.-J., Ke X., Poyet J.-L., Manji G.A.,
RA   Merriam S., Glucksmann M.A., DiStefano P.S., Alnemri E.S., Bertin J.;
RT   "CARD10 is a novel caspase recruitment domain/membrane-associated
RT   guanylate kinase family member that interacts with Bcl10 and activates
RT   NF-kappa B.";
RL   J. Biol. Chem. 276:21405-21409(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   PubMed=11356195;
RA   Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RT   "Carma1, a CARD-containing binding partner of Bcl10, induces Bcl10
RT   phosphorylation and NF-kappaB activation.";
RL   FEBS Lett. 496:121-127(2001).
RN   [3]
RP   ERRATUM.
RA   Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RL   FEBS Lett. 505:198-198(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
```

```
FT   CONFLICT     289    289       Q -> R (IN REF. 4).
FT   CONFLICT     917    917       K -> KQ (IN REF. 4).   CAB63075).
FT   CONFLICT     932    932       R -> L (IN REF. 4).
SQ   SEQUENCE   1032 AA;  115946 MW;  8377319AB82A0949 CRC64;
```

```
Query Match                 17.2%;  Score 887.5;  DB 1;  Length 1032;
Best Local Similarity       29.2%;  Pred. No. 3.1e-37;
Matches  317;  Conservative 161;  Mismatches 419;  Indels 189;  Gaps 31;

Qy    15  DEETLWEMMESHRHIRVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL  74
          :||: ||::|::|: || |  |||||||||| :|  ||||||:
Db    23  EEDALWERIEGVRHRLARALNPAKLTPYLRQCRVIDEQDEEVLSTYRFPCRVNRTGRLM  82

Qy    75  DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPVDFSNFSGLMETSKLTECLAGAIGS  134
          |:|  RGK: | AFLE:L:F:NP: :TL:TG :P :  :  :::  :   |:  :
Db    83  DILRCRGKRGYEAFLEALEFYYPEHFTLLTGQEPAQRCSMILDEEGPEGLTQFLMTEVRR  142

Qy   135  LQEELNQEKGGKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRL  194
          |:|  :Q: :  :VL   |Q:LQ:||   |   :|:: : ||          |      |
Db   143  LREARKSQLOREQQLQARGRVLEEERAGLEQRLRDQQQAQERCQRLREDWEAGSLELLRL  202

Qy   195  KDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEEQSLRTA  254
          KDE  :::::  : |:|  :  : ::  |:    :::::|:::      |||: ----LRRA
Db   203  KDENYMIAMRLAQLSEEKNSAVLRSRDLQLAVDQLKLKVSR------LEEECAL-----LRRA  254

Qy   255  SDQESDEELNRLKEENE-----------KLRS----LTFSLAE------  283
          : ::  EE::: |:|:             ||:     |
Db   255  RGPPPGAEEKBEKEKEKEPDNVDLVSELRAENQQLTASLRELQEGLQQEASRPGAPGSE  314

Qy   284  ---KDILEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMAC  340
             ::|: || :|  || ||::| ||:   |  || ::|: :  ::  :  :  :
Db   315  RILLDILEHDWREAQDSRQELCQKLHAVQGELQWAEELRDQYLQEMEDLRLKHRTLQKDC  374

Qy   341  QLYREKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELR  400
          :||:: : |: |:: |:||||||| :|||:: :  :|||:||:  ::|:  :| : |:
Db   375  DLYKHRMATVLAQLEIEKERDQAIQSRDRIQLQYSQSLIEKDQYRKQVRGLEAERDELL  434

Qy   401  TQLRLQAEPPGVLKQEARTR-EPCPREKQRLVRMHAICPRDDSDCSLVSST----  451
          T L  L    :L::::  ::       : ::
Db   435  TTLTSLEGTKALLEVQLQRAQGGTCLK----------ACASSHSLCSNLSSTWSLSEFPS  484

Qy   452  --------ESQLLSDLSATSSRELVDSFR------SSSPAPPSQQSLYKRVAEDFGEEP-  496
                  | ::   :: :   |:|           :| :::::   :
Db   485  PLGGPEATGEAVMGGPEPHNSEEATDSEKEINRLSILPFPSAGSILRRQREEDPAPPK  544

Qy   497  WSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVL  556
          |SFSS :                                |:     |
Db   545  RSFSSMSDI--------TGSV---------------------TLKPWSPGLSSSSSSDSV  575

Qy   557  MRRRPARRILSQVTMLAFQGDALL-----------EQISVIGGNLTGIFIHR  597
          : ::   :|                        :|
Db   576  WPLGKPEGLLARGCGLDFLNRSLAIRVSGRSPPGGPEPQDKGPDGLSFYGDRWSGAVVRR  635

Qy   598  VT--PGSA----ADQMALRPGTQIVMVDYEASEPLFKAVL--EDTTLEEAVGLLRRVDGF  649
          |    |||    :|:  :  :: :       |   |
Db   636  VLSGPGSARMEPREQRVEAAGLEGACLEAEAQQ--RTLLWNQGSTLPSL-----MDSK  686

Qy   650  CCLSVKVNTDGYKRLLQDLEAKVATSG-DSFYIRVNLAMEGRAKGE-LQVHCNEVLHVTD  707
          |                          ::  :  ||   : |:     :
Db   687  ACQSFH----------------------EALEAWAKGPGAEPFYIRANLTLPERADPHALCVKAQEILRLVD  736

Qy   708  TMFQGCGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQOLIALIQDMTQQCTVTRKPSS--G  765
          :                          ||| :RAQQ L    : :: |        :
Db   737  SAYKRRQEWFCTRVDPLTLRD-LDRGTVPNYQRAQQLL----EVQEKCL----PSSRHR  786

Qy   766  GPQKLVRIVSMD-------KAKASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTL  818
          ||   : :                     :|  ||    ::     : ::    :|  :|
Db   787  GPRSNLKKRALDQLRLVRPKPVGAPAGDSPDQLLLEP------C--AEPERSLRPYSL  836

Qy   819  VWPHRPARPRPVLLVPRAVGKILSEKLCLLQG----FKKCLAEYLSQEEYEAWSQRGDII  874
          V    |: ||:| | | :  :|: :|| | |     |: | ||  |  ||   |
Db   837  VRPLLVSALRPVVLLPECLAPRLIRNLLDLPSSRLDFQVCPAESLSGEELCPSSAPGAPK  896
```

```
RA   Bagguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA   Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA   Burton J., Carder C., Carter C., Chen Y., Clark G., Clegg S.M.,
RA   Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA   Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
RA   Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA   Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA   Grafham D., Griffiths M.N., Hall R., Hall-Tamlyn G.,
RA   Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA   Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA   Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA   Matthews L., Mccann O.T., Mcclay J., Mclaren S., Mcmurray A.A.,
RA   Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA   Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA   Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA   Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA   Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA   Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA   Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA   Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA   Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA   Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA   Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA   Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,
RA   Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA   Sloan D., Song L., Wang J., Wang Y., Wang Z., Murray J., Willingham D.,
RA   Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Bentley D., Bradshaw H.,
RA   Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Graves T.,
RA   Bourne S., Cordes M., Du Z., Fulton L., Goela D., Layman D., Ozersky P.,
RA   Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Pepin K.,
RA   Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,
RA   Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA   Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA   Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA   Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA   Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA   O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA   Khan A.S., Lane L., Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
CC   -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC   -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC       interaction. They both participate in a complex with MALT1, where
CC       MALT1 binds to Bcl10 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- TISSUE SPECIFICITY: Detected in adult heart, kidney and liver;
CC       lower levels in intestine, placenta, muscle and lung. Also found
CC       in fetal lung, liver and kidney.
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC       like domain. But none of these 3 domains are detected by PROSITE,
CC       Pfam or SMART.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to various
CC       gene identification problems.
```

```
DR   EMBL; AY028896; AAK26165.1; -.
DR   EMBL; AY032928; AAK54454.1; -.
DR   EMBL; AL049851; CAB63075.1; ALT_SEQ.
DR   EMBL; AL049851; CAB63076.1; ALT_SEQ.
DR   EMBL; AL022315; CAB42832.1; ALT_SEQ.
DR   PROSITE; PS50209; CARD; 1.
KW   Coiled coil.
FT   DOMAIN       23    115       CARD.
FT   DOMAIN      138    456       COILED COIL (POTENTIAL).
FT   DOMAIN      567    574       POLY-SER.
```

```
Qy      75  DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGS  134
            |:|: ||| | |||| :| | :  : :||||:| ::
Db      83  DILRCRGKRGFEAFLEALEFYYPEHFTLLTGQEPAQRCSMILDEEGPGLTQFLMTEVRR  142

Qy     135  LQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRL  194
            :|:  | ||| |  |  | :  :||:|  |||  :|||
Db     143  LREARKSQLHREQQLQARGRALEEERAGLEQRLREQQQAQERCORLREDWEAGSLELLRL  202

Qy     195  KDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELEQEQSLRTA  254
            |||                |:| |  | |  |  :  ||
Db     203  KDENYMIAMRLAQLSEEKNSAVLRSRDLQLAVDQLKLKVSR----LRRA  254

Qy     255  SDQESGDEE----------LNRLKEENEKLRSLTFSLAE  286
            :| ||  |:  ||:| |::|
Db     255  RGPPPGAEEKEREPDGADLLSELRAENQRLTASLQELQEGLQQEMSRPGAAGSERILLDI  314

Qy     287  LEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREK  346
            || |:  |:||| | :|| :|:| | | || || |:||||
Db     315  LEHDWREAQDSRQELCQKLHAVQGELQWAEELRDKYLQEMEDLRLKHRTLLKDCDLYKHR  374

Qy     347  VNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDLSRRQVFELTDQVCELRTQLRQL  406
            : :|:  ||::|||||| || ::| :|||||:| :| ||| |:|
Db     375  MATVLAQLEEIEKERDQAIQSRDRIQLQYSQSLIEKDQYRKQVRGLEAERDELLTTVTSL  434

Qy     407  QAEPPGVLKQEARTR-EPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDL------  459
            |:  | ||: :      ::|
Db     435  EGTKAMLEAQLQRTQGGSCLK----------ACASSHSLCSNLSSTWS--LSEFPSPLGG  482

Qy     460  -SAT---------SSRELVDSFR----------SSSPAPPSQQSLYKRVAEDFGEEP-WSFSSCL  503
            :|  |: :  : :|  |  ||||
Db     483  PEATGEAGGSEPHTSEEATDSEKEINRLSILPFPSAGSILRRQREEDEPPKRSFSSMS  542

Qy     504  EIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRRPAR  563
            :|  ||| :| |  |:     |:
Db     543  DI----TGSV---------------TLKPWSPGLSSSSSSDSVWPLGKPE  573

Qy     564  RILSQVTMLAFQGDALLEQIS------------VIGGNLTGIFPIHRVT--PGS  602
            :|       || ||  |::
Db     574  GLLARGCGLDFLNRSLAIRVSGWSPPAGLDPQDKSPDSMPGLGDRWSGAVVRRVLSGPGS  633

Qy     603  A-ADQMALR-PGTQIVMVDYEASEPLFKAVLEDTTL--EEAVGLLRRVDGFCCLSVKVNT  658
            | :|||  :   |  |    ||                 |||:
Db     634  ARTEQKEPRAEGTGLEGAGLEAE------AQQRTLPWNQSSTLPFLLDSKACHSFHEAL  686

Qy     659  DGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGE-LQVHCNEVLHVTDTMFQGCGCWH  717
            :|  |  ||:|  |    :| | ||   |: :    |:
Db     687  DAWAK---------GPGAEPFYIRANFSLPERSDPHALCVKAQEIILRLVDPAHKRRQEWF  737

Qy     718  AHRVNSYTMKDTAAHGTIPNYSRAQQOLIALIQDMTQQCTVTRKPSSGGPQKLVRIV--  774
            || :| |||:  ||  |||| || :| |||:|  || :   ::||:
Db     738  CTRVDTLTLRD-LDRGTVPNYQRAQQLL-EVQEKYLISSRHRSPRSNLKKRALGLVRPK  794

Qy     775  -----SMDKAKASPLRLSFDRGQLDPSRMEGSSTCFWAESC---LTLVPTLVWPHRPA  825
            :| | | |:|  |   |   |
Db     795  PAGGTAGDSAEQLP----------------AEPCSELERSLKPYSLVRPLLVS  831

Qy     826  RPRPVLLVPRAVGKILSEKLCLLQG---FKKCLAEYLSQEY-----EAWSQRGD  872
            ||:::| |:|  |  :|  | |||     : |  |: :|           |: |
Db     832  ALRPVVLLPECLAPRLJRNLLDLPSSRLDFQVCPAESLSGEEQCTSSAPGAPKAWPATAG  891

Qy     873  IIQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMA  932
            :|:|| |:|| ||| |:|| ||:|| |||||||| :|| | || |||||||||:||||:
Db     892  L------GSR--IRAIQESVGKK--HCLLELGARGVRELVHSEVYPIVIHVEVTEKNV  939

Qy     933  KKLKKGLQRLGTSEEQLLEAARQEEGDLDRAPCLYSSLAPDGWSDLDGLLSCVRQAIADE  992
            ::::  :::| ||| ||| | :||| || | |||| :|| |||  :| |||| ||:: |
Db     940  REIRGLLGRPGWRDSELLRQCRGSEQWLWGLPCSWVQVPAHAWGHAEELAKVVRGRILQE  999

Qy     993  QKKVVWTEQ 1001
            : :|||
Db    1000  QARLVWVER 1008
```

```
RESULT 5
CARA_MOUSE
ID   CARA_MOUSE            STANDARD;      PRT;   1021 AA.
AC   P58660;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Caspase recruitment domain protein 10 (Bcl10-interacting MAGUK protein
DE   1) (Bimp1).
GN   CARD10 OR BIMP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21391892; PubMed=11387339;
RA   McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA   Li Y., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA   Nunez G.;
RT   "Bimp1, a MAGUK family member linking protein kinase c activation to
RT   Bcl10-mediated NF-kappa B induction.";
RL   J. Biol. Chem. 276:30589-30597(2001).
CC   -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC   -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC       interaction. They both participate in a complex with MALT1, where
CC       MALT1 binds to Bcl10.
CC   -!- TISSUE SPECIFICITY: Highly expressed in kidney, heart followed by
CC       brain, lung, liver, skeletal muscle and testis.
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC       like domain. But none of these 3 domains are detected by PROSITE,
CC       Pfam or SMART.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF363456; AAK60136.1; -.
DR   PROSITE; PS50209; CARD; 1.
KW   Coiled coil.
FT   DOMAIN        23    115       CARD.
FT   DOMAIN       138    450       COILED COIL (POTENTIAL).
FT   DOMAIN       558    565       POLY-SER.
SQ   SEQUENCE   1021 AA;  114413 MW;  481lA09BDB8F792C CRC64;
```

```
Query Match             17.2%;  Score 886.5;  DB 1;  Length 1021;
Best Local Similarity   28.8%;  Pred. No. 3.5e-37;
Matches  314;  Conservative 156;  Mismatches 414;  Indels 205;  Gaps 31;

Qy      15  DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRLTNSAMRAGHLL  74
            :  |:::  ::  :  |  :|||: ||| |:|  : |::  ||||
Db      23  EEDALWERIEGVRHRLTRALNPAKITPYLRQCRVLDEQDEEEVLSTYRFPCRANRTGRLI  82
```

```
RESULT 6
CAR9_RAT
ID   CAR9_RAT     STANDARD;      PRT;    536 AA.
AC   Q9EPY0;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Caspase recruitment domain protein 9 (rCARD9).
GN   CARD9.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20576268; PubMed=11053425;
RA   Bertin J., Guo Y., Wang L., Srinivasula S.M., Du M.-Q., Dyer M.J.S., Robison K.E.,
RA   Poyet J.-L., Merriam S., DiStefano P.S., Alnemri E.S.;
RT   "CARD9 is a novel caspase recruitment domain-containing protein that
RT   interacts with Bcl10/CLAP and activates NF-kappa B.";
RL   J. Biol. Chem. 275:41082-41086(2000).
CC   -!- FUNCTION: Activates NF-kappaB via Bcl10 (By similarity).
CC   -!- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by
CC       CARD-CARD interaction (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AF311288; AAG28791.1; -.
DR   InterPro; IPR001315; CARD.
DR   PROSITE; PS50209; CARD; 1.
KW   Coiled coil.
FT   DOMAIN        6     98       CARD.
FT   DOMAIN      117    277       COILED COIL (POTENTIAL).
FT   DOMAIN      303    420       COILED COIL (POTENTIAL).
SQ   SEQUENCE   536 AA;   62631 MW;  6F33089CB7E6BAC9 CRC64;

Query Match             10.0%;  Score 512.5;  DB 1;  Length 536;
Best Local Similarity   28.8%;  Pred. No. 7.3e-19;
Matches  162;  Conservative   94;  Mismatches 215;  Indels  91;  Gaps  15;

Qy   15 DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL 74
             :: :|| ||||||:||||| ||  |||:|| |||   ||:|: | ||
Db    6 NDDECWSALESFRVKLISVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRRKKVGVLL 65

Qy   75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPD-------VDFSNFSGL---------ME 120
           |:|:|   |:||||||||| ||||:||:|| || |   ||:|| |:|
Db   66 DLIQRTGHGKGVYAFLESLELYPQLYRKVTGKEPARVFSMIIDASGESGLTQLLMTEVMK 125

Qy  121 TSKLTECLAGAIGSLQEELNQERGQKEVLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
          |:  |:  | |:| ||:| :||: || |   :|
Db  126 LQKKVQDLTALLSS-KDDFIKELRVKDSLLRKHQERVQ--------------------RL 164

Qy  181 KREVSAHFHEVLRIKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
           |: |   | | |:   |:| |   |  ||::|:  |:   ||:  |
Db  165 KEECELSSAELKRCKDENYDLAMRLAHLSEEKGAALMRNRDLQLEVDQLRHSLMKAE-D 222

Qy  241 SCELELQEQ-SLRTASDQESGDEELNRLKEENE---KLRSLTFSLAEK--------DI 286
           | |:|:|    ||:| :|:|| ::|::|  |   |:|  ||:|
Db  223 DCKVERKHTLKLRHAMEQRPSQELLWDLQQERDLLQARVQELEVSVQEGKLHRNSPYIQV 282

Qy  287 LEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREK 346
           ||  |||  || |::| |::: |  |  |:  | ||||  |||
```

```
Db  283 LEEDWRQALQEHQEQASTIFSLRKDLRQAEALRTCMEEKEMFELQCLALRKDAKMYKDR 342

Qy  347 VNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQL 406
            :| |::  :: | |||||  ||| | |:  |:| :||||  |: | ||||| || |||
Db  343 IEATLQQMEEVSIERDQAMTSREELHAQCAQSFQDKDKLRKQVRELDEKADELQLQLFQT 402

Qy  407 QAEPPGVLKQEARTREPCPREKQRLV---RMHAICPRDDSDCSLVSSTESQL-LSDLSAT 462
           | | |  || ||  |   |  :|| |    |  |   || :    ||: |  |||:|||
Db  403 ESR---LLAAEGRLKQ--QQLDMLLSSDLEDSSPPNSQELSLPQDLEEDAQLSDKGVL 456

Qy  463 SSRE---------LVDSFRSSSPAPPSQQSLYKRVAEDFG--------EEPWS 498
           |||             :        ||  |::                  :
Db  457 ADRESPEQPFVVLNKKHLSQTHDTVPSSSEPPEKERRRLKESFENYRRKRALRKMQNSWR 516

Qy  499 FSSCLEIPEGDPGALPGAKAGD 520
           ||          ::
Db  517 QG------EGDHGNTTGSDNTD 532


RESULT 7
CAR9_HUMAN
ID   CAR9_HUMAN     STANDARD;      PRT;    536 AA.
AC   Q9H257; Q9H854;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Caspase recruitment domain protein 9 (hCARD9).
GN   CARD9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20576268; PubMed=11053425;
RA   Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA   Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E.,
RA   DiStefano P.S., Alnemri E.S.;
RT   "CARD9 is a novel caspase recruitment domain-containing protein that
RT   interacts with Bcl10/CLAP and activates NF-kappa B.";
RL   J. Biol. Chem. 275:41082-41086(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retinoblastoma;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba S., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-492 FROM N.A.
RC   TISSUE=Muscle;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Activates NF-kappaB via Bcl10.
CC   -!- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by
CC       CARD-CARD interaction.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Highly expressed in spleen. Also detected in
CC       liver, placenta, lung, peripheral blood leukocytes and in brain.
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   -!- CAUTION: Ref 2 sequence differs from that shown due to a
CC       frameshift in position 360.
CC   ---------------------------------------------------------------------
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=93529086; PubMed=8486731;
RA   Itoh M., Nagafuchi A., Yonemura S., Yasuda-Kitani T., Tsukita S.,
RA   Tsukita S.;
RT   "The 220-kD protein colocalizing with cadherins in non-epithelial
RT   cells is identical to ZO-1, a tight junction-associated protein in
RT   epithelial cells: cDNA cloning and immunoelectron microscopy.";
RL   J. Cell Biol. 121:491-502(1993).
CC   -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
CC       REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY
CC       HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
CC       MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.
CC   -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
CC       MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
CC       OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN
CC   -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D14340; BAA03274.1; -.
DR   PIR; A46431; A46431.
DR   HSSP; P31016; 1BE9.
DR   MGD; MGI:98759; Tjp1.
DR   InterPro; IPR000619; Guanylate_kin.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR009906; ZU5.
DR   Pfam; PF00625; Guanylate_kin; 1.
DR   Pfam; PF00595; PDZ; 3.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 3.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 3.
DR   PROSITE; PS50002; SH3; 1.
KW   Tight junction; SH3 domain; Repeat; Membrane.
FT   DOMAIN       23    110       PDZ 1.
FT   DOMAIN      186    264       PDZ 2.
FT   DOMAIN      421    502       PDZ 3.
FT   DOMAIN      516    584       SH3.
FT   DOMAIN      644    794       GUANYLATE KINASE.
FT   DOMAIN     1242   1247       POLY-PRO.
FT   DOMAIN     1424   1430       POLY-PRO.
SQ   SEQUENCE   1745 AA;  194710 MW;  C3DA2C0A9F411F66 CRC64;

Query Match             5.6%;  Score 289;  DB 1;  Length 1745;
Best Local Similarity   22.3%;  Pred. No. 4.5e-07;
Matches  191;  Conservative  136;  Mismatches  319;  Indels  212;  Gaps  43;

Qy    231 QELQRANMVSSCELELQEQSLRTASDQESGDEELNRLKEENKLRSLTFSLAEKDILEQS 290
              || ||:  ::  :| : :::||   |||||
Db     70 QENDRVAMVNGVSMDNVEHAFAVQQLRKSGKNA--KITIRRKKKVQIPVSHPDPEPVSDN 127

Qy    291 LDEARGSRQELVERIHSLRE-RAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNA 349
                                :| ::  ||:::|          :::
Db    128 EDDSYD------EEVHDPRAGRGALANRRSEKSWARDRSA---SRERSLSPRSDRRSVAS 178
```

```
RT   tight as well as adherens junctions with a binding affinity to
RT   occludin and alpha catenin.";
RL   J. Biol. Chem. 274:5981-5986(1999).
CC   -!- FUNCTION: PLAYS A ROLE IN TIGHT JUNCTIONS AND ADHERENS JUNCTIONS.
CC   -!- SUBUNIT: INTERACTS WITH OCCLUDIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
CC   -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF113005; AAD19964.1; -.
DR   HSSP; P31016; 1BE9.
DR   MGD; MGI:1341872; Tjp2.
DR   InterPro; IPR000619; Guanylate_kin.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00625; Guanylate_kin; 1.
DR   Pfam; PF00595; PDZ; 3.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 3.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 3.
DR   PROSITE; PS50002; SH3; 1.
KW   Tight junction; SH3 domain; Repeat; Membrane.
FT   DOMAIN        10     97       PDZ 1.
FT   DOMAIN       287    365       PDZ 2.
FT   DOMAIN       489    570       PDZ 3.
FT   DOMAIN       584    649       SH3.
FT   DOMAIN       678    858       GUANYLATE KINASE.
FT   DOMAIN      1139   1142       POLY-GLU.
SQ   SEQUENCE   1167 AA;  131614 MW;  F15DA3EBC3F9434F CRC64;
```

```
Query Match             5.4%;  Score 280;  DB 1;  Length 1167;
Best Local Similarity   23.0%;  Pred. No. 7.7e-07;
Matches  198;  Conservative 122;  Mismatches 308;  Indels 234;  Gaps  43;
```

```
Qy   248  EQSLRTASDQESGDEELNR--------------LKEENEKLRSLTFSLAEKDI----------  286
          |:|| :  ||  :: |  ||              : || ||| ||| ::|
Db   134  ERSRHSSHDMLSHSWEGNRERGGPHQRTQSRERSRGRSLERGLDQEDYGRSRERSRGR  193

Qy   287  -LEQSLDEARGSRQELVERIHSLRERAVAAERQR------------------EQY-------  327
           ||:|          :|| ||| :||:||:|  |                  :::
Db   194  SLERGLD------RDFVSRDHS-RGRSIDRDYRDYERSYHEAYEPDYGGGYSPSYDRRA  246

Qy   328  QTLLQFQKSKMACQLYREKVNALQAQ-------------------VCELQKERDQAYSARDSA  371
          ::|| |:|  | |  : :|| || |:                   ||::||: :||:|
Db   247  HPETRYERSRSREHL-RSRSPSPESRSRREHKGQHDPRPIGVLLTQSKANEEVGLRLGS  305

Qy   372  Q---REISQS-LVEKDSLRRQ---VFELTDQVCE--LRTQLRQLQAEPPGVLKQEARTRE  422
          |   ||: |  |:|||||||      :: |  : |  ::|||:: | | |: |:  ::|
Db   306  QIFIKEMTRTGLATKDGNLHEGDIILKINGTVTENMSLTDARKLIEKSRGKL-QLVVLRD  364

Qy   423  PCPREKQRLVRMHAICPRDD--SDCSLVSSTES-----QLLSDLSATSSRELVDSFRS  473
          | | :|  :|: :|| |  |   :| :|::|:|      ||:|||::::|||:|||:
Db   365  ----SKQTLNIPALNDRDSEVEDISEIESNRSFSPEERRQQYSDQDYHSSTEKL-----  415

Qy   474  SSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEI-PEGPDGALPGAKAGPHLDYELLDTAD  532
          :  :| : ||| |  | || :::  |:: ::|  |:|| ||| : | |||: |
Db   416  -KERPSSREETSGRLSR-MGATPTPFKSTGDITPAGVTEA----TREPRYQEE----GP  464
```

```
Qy   350  LQ---AQVCELQKERDQAYSARDSAQ---REISQ-SLVEKDSLRRQ---VFELTDQVCE   398
          :|     ::  ||: :|:  | :::|   : | |  ||||| ||||      :|  |  |
Db   179  SQPAKPTKVTLVKSRKNEEYGLRPASHIFVKEISQDSLAARDGDIQEGDVVLKINGTVTE   238

Qy   399  --LRTQLRQLQAEPPGVLKQEARTREPCCPREKQRLVR------MHAICPRDDSDCSLVS   449
            :||:|  | | |:  | :|:   |: |                :  :| ::| ::| :
Db   239  NMSLTDAKTLIERSKGKLKMVVQR-----DERATLLNVPDLSDSIHSANASERDDISEIQ   293

Qy   450  STESQLLSDLSATS-------------SRELVDSFRSSSPAPPSQQSLYKRVAEDFGEE   495
          ||:| ||:|||:|              ||  : | :: | | |::| |:| : |
Db   294  S----LASDHSGRSHDRPPRRSQSRSPQQRSEPSDHSTQSPQQPSNGSLRSR-EERRMSK   348

Qy   496  PWSFSSCLE--------------------IPEGDPGALPGAKAGDPHLDYELLDTA   531
          |  | |                       :||:| ||| : | |||: |||:||
Db   349  PGAISTPVKHVDDHPPKAVEEVTVEKNEKQTPTLPEPKP---VYAQVGQPDVDL------   399

Qy   532  DLPQLESSLQPVSP--GRLDVS--ESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIG   587
          ||| |:|| |     | |  :   |:|:| :    | |: | :: | |:|  ||:| :|
Db   400  --------PVSPSDGALPNSAHEDGIL---RPSMKLVK-----FRKGDSV--GLRLAG   439

Qy   588  GNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTL-------EEA   639
          || |:| :| | ||:|  |   | | |::| |:|:| | |:  | :::         |:
Db   440  GNDVGIFVAGVLEDSPAAKEGLEEGDQILRVN---NVDFTNIIREEAVLFLLDLPKGEEV   496

Qy   640  VGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHC   699
          ||||| |||   |:|  ||||  ||  ||:||:|:|||||: |:  :||: |:||||  |
Db   497  TILAQK-----------KKDVYRRIVE-----SDVGDSFYIRTHFEYEKESPYGLSFNK   539

Qy   700  NEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQLIALIQDMTQQCTV   758
          || | | |:| |  |   |   |   : |::: | |||: |||:| | | ||:||| ||
Db   540  GEVFRVVDTLYNGKLGSWLAIRIGKNHKE--VERGIIPNKNRAEQ--LASVQ-------Y   588

Qy   759  TRKPSSGGPQKLVRIVSMDKAKASPLR-LSFDRGQLDPSRMEGSSTCFWAESCLTLVP-Y   816
          ||:| |  |:|  |:| ||:|||: |  |    ||:|   :|    | |||| ::|  |
Db   589  TLPKTAGG---------DRADEWRFGLRSSKRNLRKSREDLS------AQPVQTKFPAY   633

Qy   817  TLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEE--YE-AWSQRGD   872
          || | |   |   |    |:|: | | | || ||:   | :|| |    || | |: |
Db   634  ERVVLREAGFLRPVTIF-GPIADVAREK---------LAREEPDIYQIAKSELRD   678

Qy   873  IIQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMA   932
          ::|:|   |    | : : |  :::::| ||||  || :|:::  | ||  :| |: |
Db   679  AGTDHRSSG---IIRLHTIKQIIDQDKHALLDVTPNAVDRLNYAQWYPIVVFLNPDSKQG   735

Qy   933  KKL--------KKGLQRLGTSEEQLLEAARQEEGDLDRAPCLYSSLAPDGWSDLDGLLS   983
          |::        |||  |  :::|:| :|:  ||:| |  | |::|::|| |:||:| |
Db   736  VKTMRRLCPESRKSARKLYERSHKL----RKNNHLFTTINLNSM-NDGW------YG   784

Qy   984  CVRQAIADEQKKVVWTEQ 1001
          : :: |::||:: |  ||
Db   785  ALKEAIQQQNQLWVSE   802
```

```
RESULT   9
ID   ZO2_MOUSE     STANDARD;     PRT;   1167 AA.
AC   Q9Z0U1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tight junction protein ZO-2 (Zonula occludens 2 protein) (Zona
DE   occludens 2 protein) (Tight junction protein 2).
GN   TJP2 OR ZO2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C56BL/6 X CBA;
RX   MEDLINE=99150392; PubMed=10026224;
RA   Itoh M., Morita K., Tsukita S.;
RT   "Characterization of ZO-2 as a MAGUK family member associated with
```

```
Qy   533  LPQLESSLQPVSPGRLDVSESGVLMRRRPARRIL--SQVTMLAF-QGDALLEQISVIGGN  589
          :||    || :    :: |  |  ||:|  :||   :  :||:
Db   465  VP------QPRAAPR------VFLRPSPDEAIYGPNTKMVRFKKGDSV--GLRLPGGN   509

Qy   590  LTGIFIFHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTL------EEAVGL  642
          |||  :  || :|| ::|  |  |:  |:  ::: | :   ||
Db   510  DVGIFVAGIQEGTSAEQEGLQEGDQILKVNTQDFRGL--VREDAVLYLLEIPKGETVTI   566

Qy   643  LRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEV  702
          |  :  :   : :   |       ||||:|| : |
Db   567  L------AQSRADVYRDIL------ACGRGDSFFIRSHFECEKETPQSLAFTRGEV   610

Qy   703  LHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCT----  757
          | ||::  :  ||:|  |  |||    |   |||||:||:
Db   611  FRVVDTLYDGKLGHWLAVRIGNDLEK------GLIPNKSRAEQ------MDSVQNAQRENA  659

Qy   758  ------VTRKPSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSSTCFWAES  809
                :: |||| : |:  ::  |  | ::  :| :|:|
Db   660  GDRAVFWVMRRQRSRGGDKKTLR------------KSREDLARSVSVSTKFPA--   700

Qy   810  CLTLVPYTLVWPHRPARPRPVLL---VPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYE  865
          ||:|  |   ||    |:  |::      |:||
Db   701  ------YEKVLLREAGFKRPVLFGPIADIAMERLANELPDLFQTAK-----TEPKD  746

Qy   866  AWSQRGDIIQEGEVSGGRCWTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHV  925
          |     : :  :      : | :  |  ||:|||:|||:  :||  |:|:| |:|:|
Db   747  AGSEKSSGV----------VRLNTVRQIIEQDKHALLDVVTPKAVDLLNYTQWFPIVIFF  795

Qy   926  SVNEKMAKKLKKGLQRL----GTSEEQLEAARQEEGDLDR--APCLYSSLAPDGWSDLD  979
          | :    || |  ||      | |  :| :  |:|  || :          || |:|:|
Db   796  NPDSRQGVKTIR--QRLSPTSNKSSRKLFDQANKLKKTCSHLFTATINVNSANDGW---  849

Qy   980  GLLSCVRQAIADEQKKVVWTEQ 1001
          | : |   |    :   ||
Db   850  --FGSLKDSIQQQNEAVWVSE 869
```

```
RESULT    10
ZO1_HUMAN
ID    ZO1_HUMAN    STANDARD;    PRT;    1736 AA.
AC    Q07157;
DT    01-OCT-1994 (Rel. 30, Created)
DT    01-OCT-1994 (Rel. 30, Last sequence update)
DT    16-OCT-2001 (Rel. 40, Last annotation update)
DE    Tight junction protein ZO-1 (Zonula occludens 1 protein) (Zona
DE    occludens 1 protein) (Tight junction protein 1).
GN    TJP1 OR ZO1.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=Liver;
RX    MEDLINE=93361541; PubMed=8395056;
RA    Willott E., Balda M.S., Fanning A.S., Jameson B., van Itallie C.,
RA    Anderson J.M.;
RT    "The tight junction protein zo-1 is homologous to the Drosophila
RT    discs-large tumor suppressor protein of septate junctions.";
RL    Proc. Natl. Acad. Sci. U.S.A. 90:7834-7838(1993).
CC    -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
CC        REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY
CC        HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
CC        MIGHT BE INVOLVED IN STABILIZING TIGHT JUNCTIONS.
CC    -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.
CC    -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
CC        MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
CC        OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT.
CC    -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC        SHORT FORM; ISOFORMS ARE PRODUCED BY ALTERNATIVE SPLICING.
CC    -!- TISSUE SPECIFICITY: THE ALPHA-CONTAINING ISOFORM IS FOUND IN MOST
CC        EPITHELIAL CELL JUNCTIONS. THE SHORT ISOFORM IS FOUND BOTH IN
CC        ENDOTHELIAL CELLS AND THE HIGHLY SPECIALIZED EPITHELIAL JUNCTIONS
CC        OF RENAL GLOMERULI AND SERTOLI CELLS OF THE SEMINIFEROUS TUBULES.
CC    -!- PTM: PHOSPHORYLATED.
CC    -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC    -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC    -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC    -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; L14837; AAA02891.1; -.
DR    PIR; A47747; A47747.
DR    HSSP; P31016; 1BFE.
DR    MIM; 601009; -.
DR    InterPro; IPR000619; Guanylate_kin.
DR    InterPro; IPR001478; PDZ.
DR    InterPro; IPR001452; SH3.
DR    InterPro; IPR000906; ZU5.
DR    Pfam; PF00625; Guanylate_kin; 1.
DR    Pfam; PF00595; PDZ; 3.
DR    Pfam; PF00018; SH3; 1.
DR    Pfam; PF00791; ZU5; 1.
DR    SMART; SM00072; GuKc; 1.
DR    SMART; SM00228; PDZ; 3.
DR    SMART; SM00326; SH3; 1.
DR    SMART; SM00218; ZU5; 1.
DR    PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR    PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR    PROSITE; PS50106; PDZ; 3.
DR    PROSITE; PS50002; SH3; 1.
KW    Tight junction; SH3 domain; Alternative splicing; Repeat; Membrane;
KW    Phosphorylation.
FT    DOMAIN      11     98       PDZ 1.
FT    DOMAIN     174    252       PDZ 2.
FT    DOMAIN     409    490       PDZ 3.
FT    DOMAIN     504    572       SH3.
FT    DOMAIN     632    782       GUANYLATE KINASE.
FT    DOMAIN    1231   1236       POLY-PRO.
FT    DOMAIN    1414   1420       POLY-PRO.
FT    VARSPLIC   910    989       MISSING (IN SHORT ISOFORM).
SQ    SEQUENCE   1736 AA;   194721 MW;   50B0D1B7A081AFFE CRC64;
```

```
Query Match             5.3%;     Score 270.5;  DB 1;  Length 1736;
Best Local Similarity  21.4%;     Pred. No. 3.8e-06;
Matches  174;  Conservative 129;  Mismatches 278;  Indels 233;  Gaps  38;

Qy   247  QEQSLRTASDQESGDEE---------LNRLKEENEKLRSLTFSLAEKDILEQSLDEARG  296
          :|:||| ||:|      ||  ||   ||   |
Db   151  RERSLSPRSDRRSVASSQPAKPTKVTLVKSRKNEEYGLR-LASHIFVKEISQDSLAARDG  209

Qy   297  SRQE--LVERIH-SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQ  353
          |:|    ||:::    |:::  || : :   :  ||:||
Db   210  NIQEGDVVLKINGTVTENMSLTD--------AKTLIERSKGKLKMVVQRDE---RAT  255

Qy   354  VCELQKERDQAYSARDSAQREIS--QSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPP  411
            : ||  :     :|  : | :    ||:    ||   |::
Db   256  LLNVPDLSDSIHSANASERDDISEIQSLASDHSGRS------------------HDRPP  296

Qy   412  GVLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQL-------LSDL  459
          | |:||| |::|  | |:||   :|   ::| :|:: :  :||
Db   297  ----RRSRSRSPDQRSEPSDHSRHS--PQQPSNGSLRSRDEERISKPGAVSTPVKHADDH  350

Qy   460  SATSSREL-VDSFRSSSPAPPSQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKA  518
          :|||::|  ||   :|    ||:  ||    |  |:    |:|
Db   351  TPKTVEEVTVERNEKQTPSLPEPKPVYAQV---GNQMWIYLSV----------------  390

Qy   519  GDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRPARRILSQVTMLAFQGDA  578
```

```
RESULT 11
ZO2_CANFA
ID   ZO2_CANFA      STANDARD;      PRT;   1174 AA.
AC   Q95168;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tight junction protein ZO-2 (Zonula occludens 2 protein) (Zona
DE   occludens 2 protein) (Tight junction protein 2).
GN   TJP2 OR ZO2.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96421547; PubMed=8824195;
RA   Beatch M., Jesaitis L.A., Gallin W., Goodenough D.A., Stevenson B.R.;
RT   "The tight junction protein ZO-2 contains three PDZ (PSD-95/Discs-
RT   Large/ZO-1) domains and an alternatively spliced region.";
RL   J. Biol. Chem. 271:25723-25726(1996).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=94179414; PubMed=8132716;
RA   Jesaitis L.A., Goodenough D.A.;
RT   "Molecular characterization and tissue distribution of ZO-2, a tight
RT   junction protein homologous to ZO-1 and the Drosophila discs-large
RT   tumor suppressor protein.";
RL   J. Cell Biol. 124:949-961(1994).
CC   -!- FUNCTION: PLAYS A ROLE IN TIGHT JUNCTIONS AND ADHERENS JUNCTIONS.
CC   -!- SUBUNIT: INTERACTS WITH OCCLUDIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; L27152; AAC37332.1; -.
DR   HSSP; P31016; 1BE9.
DR   InterPro; IPR000619; Guanylate_kin.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00625; Guanylate_kin; 1.
DR   Pfam; PF00595; PDZ; 3.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 3.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00865; GUANYLATE_KINASE_1; FALSE_NEG.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 3.
DR   PROSITE; PS50002; SH3; 1.
KW   Tight junction; SH3 domain; Repeat; Membrane; Alternative splicing.
FT   DOMAIN       10     97       PDZ 1.
FT   DOMAIN      291    369       PDZ 2.
FT   DOMAIN      493    574       PDZ 3.
FT   DOMAIN      588    653       SH3.
FT   DOMAIN      680    860       GUANYLATE KINASE.
FT   DOMAIN     1146   1149       POLY-GLU.
SQ   SEQUENCE   1174 AA;  132085 MW;  45AB836BBDDB1226 CRC64;
```

```
Query Match             5.0%;
Best Local Similarity   22.9%;
Matches  196;  Conservative  124;  Mismatches  325;  Indels  210;  Gaps  46;
                        Score 256;  DB 1;  Length 1174;
                        Pred. No. 1.2e-05;

Qy   210  QEKELAASRCRSLQEEL----YLLKQELQRANMVSSCELELQEQSLRTASDQESGDEEL  264
               |||  ||.|||   :||  :    :|.       :||        ||      ::
Db   164  QERE--RSRGRSLERGLDHDDDYRRPRERSRG-------RSLERGLDH--DDDY  206

Qy   265  NRLKEENEKLRS-----------------LTFSLAEKDILEQSLDEAR--GSRQELVERI  305
          :|||            || : ||         ::     : |:::
Db   207  GRPGERSHGMSTDRGYDRGYDRGYDRTYS-PEAEYGRRTQPDARHAGSRSR--SRE  263

Qy   306  HSLRERAVAE-RQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQA  364
          | | |  |    :||                          ::|
Db   264  H-LRSRSPSPELRGRPDHAGQPD-------SDRPIGVLLMKSKAN-------EE  302

Qy   365  YSARDSAQREISQ---SLVEKDSLRRQ--VFELTDQVCE-LRTQLRQLQAEPPGVLK  415
          |     :|||||  |||  :: :|     |||            :.   :.|
Db   303  YGLRLGSQIFIKQMTRTALATKDGNLHEGDIILKINGTVTENMSLTDARKLIEKSRGKL-  361

Qy   416  QEARTREPCPREKQRLVRMHAICPRDSDSCSL-VSSTES------QLLSDLSATS  463
          |       |   :||          :|:|  ::|:    ||     :||:||:: |
Db   362  QLVLRD----SKQTLINIPSL--NDSDSEIEDISEIESNRSFSPEERRQQYSDYDYHS  414

Qy   464  SRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAK-AGDPH  522
          | : |:                  :| :    :  :: .:             ::.|:
Db   415  SNEKL----KERPNSREDMQNRWSR-MGATPTPFKSM-----GDIASVVGTENSKEPR  462

Qy   523  LDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRRPARRIL--SQVTMLAF-QGDAL  579
          :||       || :     :: :|       |  |   |:|     :|||||   |||:|
Db   463  YQEE---------PPAPQPKAAPR-----TFLRPSPEDEAIYGPNTKMVRFKKGDSV  505

Qy   580  LEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEA  639
          ||:||| |||| ||| |.:  || |:|  |  || |:||  :  |     ::|:|
Db   506  --GLRLAGGNDVGIFVAGIQEGTSAEQGLQEGDQILKVNTQDFRGL--VREDAVLY--  558

Qy   640  VGLLRRVDG-FCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVH  698
           ||||||   :| :: :|:|||||||||||:||||||||||||||
Db   391  ---HLMVSYL-------------IQLMKMGFL---RPSMKLVK-----FRKGDS  420
```

```
Qy   579  LLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTL--  636
          :||:||| |||| |||:|.: :| : :|  |  || |:||   : |    ::|.: |
Db   421  V--GLRLAGGNDVGIFVAGVLEDSPAAKEGLEEGDQILRVN---NVDFTNIIREEAVLFL  475

Qy   637  -----EEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGR  690
               :: |  ::    :|:|      :  |   ||||  |     |||||||:
Db   476  LDLPKGEEVTILAQK---------KKDVYRRIVE-----SDVGDSFYIKTHFEYEKE  518

Qy   691  AKGELQVHCNEVLHVTDMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALI  749
          : |:||      : :.:: :     |  |  :   :: ::: |:|||.:|||:|:|:
Db   519  SPYGLSFNKGEVFRAVDTLYNGKLGSWLAIRIGKNHKE--VERGIIPNKNRAEQ--LASV  574

Qy   750  QDMTQQCTVTRKPSSGGPQKLVRIVSMDKAKASPLR-LSFDRGQLDPSRMEGSSTCFWAE  808
          |                 :| |:.:|.   .:      :  :| | ||   |::  :
Db   575  Q------YTLPKTAGG-------DRADEWFRGLRSSKRNLRKSREDLS-----AQ  612

Qy   809  SCLTLVP-YTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEE---Y  864
             |::|  :|:     : ||| :|: .|:||       |  |    |:||  : ::    |
Db   613  PVQTKFPAYERVVLREAGFLRPVTIF-GPIADVAREK-----------LAREEPDIY  657

Qy   865  E-AWSQRGDIIQEGEVSGGRCWTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVI  923
          : :    ||:::   :.:      | :: |:: |||||||  .:||   |   |:|:||
Db   658  QIAKSEPRDAGTDQRSSG--YIRLHTIKQIIDQDKHALLDVTPNAVDRLNYAQWYPIVV  714

Qy   924  HVSVNEKM----------------AKKLKKGLQRLGTSEEQLLEAARQEEGLDRAPCLY  967
              :|          |||: |: | |::: |    :|  :| :      |:
Db   715  FLNPDSKQGVKTMRMRLCPESRSAARKLYERSHKLAKNNHHLFTTT-----------IN  762

Qy   968  SSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQ  1001
          :::  ||| |: |  |:||  :.:::: ::||
Db   763  LNSMNDGW-----YGALKEAVQQQQNQLVWVSE  790
```

Db   559 --LLEIPKGEMVTILAQSRADVYRDIL------ACGRGDSFFIRSHFECEKETPQSLAFS 610
                || |||:: | :|::|:| ||:: || :|:| ::|:
Qy   699 CNEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCT 757
                || |:|: | |-|-|||-|:|:: | |::|: ::| :: :|:|
Db   611 RGEVFRVRVDTLYDGKLGHWLAVRIGNELEK----GLIPNKSRAEQ--MASVQN------ 657

Qy   758 VTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGCQLDPSRMEGSSTCFWAESCLTLVPYT 817
                :: :|| | : ||| | || || | | ||
Db   658 -AQRDNAGDRADFWRM----RGQRSGMKKNLRKSRDLTRAVSVSTKFPA------YE 704

Qy   818 LWPHRPARPRPVLL---VPRAVGKILSEKICLLQGFKKCLAEVLSQEYEAWSQRGDI 873
                ||| |: |:| : |:: | | :|:|||:|
Db   705 RVLLREAGFKRPVVLFGPIADIALEKLANELPDLFQTAK-------TEPKDAGSEKSSG 756

Qy   874 IQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAK 933
                  : |:: ||| ||||| ||| | | |:: |:|||| :::| |||||
Db   757 V---------VRLNTVRQIIEQDKHALLDVTPKAVDLLNYTQWFPIVIFFNPDSRQGV 805

Qy   934 KLKKGLQRLGTSEEQLLEAARQEEGDLDRAPCLY------SSLAPDGWSDLDGLLSCVR 986
                : : ||| || ||||||||||||||||||
Db   806 KTMR--QRLNPTSNKSSRKLYDQANKLKKT-CAHLFTATINLNSANDSW------FGSLK 856

Qy   987 QAIADEQKKVVWTEQ 1001
                || : :|: ||
Db   857 DTIQHQQGEAVWVSE 871

RESULT 12
ZO2_HUMAN
ID   ZO2_HUMAN   STANDARD;   PRT; 1190 AA.
AC   Q9UDY2; Q15883; Q9UDY1; Q9UDY0; Q99839;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tight junction protein ZO-2 (Zonula occludens 2 protein) (Zona
DE   occludens 2 protein) (Tight junction protein 2).
GN   TJP2 OR ZO2 OR X104.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM A1).
RC   TISSUE=Brain;
RX   MEDLINE=95038744; PubMed=7951235;
RA   Duclos F., Rodius F., Wrogemann K., Mandel J.L., Koenig M.;
RT   "The Friedreich ataxia region: characterization of two novel genes and
RT   reduction of the critical region to 300 kb.";
RL   Hum. Mol. Genet. 3:909-914(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS A1; C1; A2 AND C2).
RC   TISSUE=Pancreas;
RX   MEDLINE=20472048; PubMed=11018256;
RA   Chlenski A., Ketels K.V., Korovaitseva G.I., Talamonti M.S., Oyasu R.,
RA   Scarpelli D.G.;
RT   "Organization and expression of the human zo-2 gene (tjp-2) in normal
RT   and neoplastic tissues.";
RL   Biochim. Biophys. Acta 1493:319-324(2000).
RN   [3]
RP   SEQUENCE OF 1-104 FROM N.A. (ISOFORMS A1 AND C1).
RC   TISSUE=Pancreas;
RX   MEDLINE=99287578; PubMed=10360833;
RA   Chlenski A., Ketels K.V., Tsao M.-S., Talamonti M.S., Anderson M.R.,
RA   Oyasu R., Scarpelli D.G.;
RT   "Tight junction protein ZO-2 is differentially expressed in normal
RT   pancreatic ducts compared to human pancreatic adenocarcinoma.";
RL   Int. J. Cancer 82:137-144(1999).
RN   [4]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORMS A1 AND C1).
RC   TISSUE=Pancreas;
RX   MEDLINE=99426875; PubMed=10495427;
RA   Chlenski A., Ketels K.V., Engeriser J.L., Talamonti M.S., Tsao M.-S.,
RA   Koutnikova H., Oyasu R., Scarpelli D.G.;

RT   "zo-2 gene alternative promoters in normal and neoplastic human
RT   pancreatic duct cells.";
RL   Int. J. Cancer 83:349-358(1999).
RN   [5]
RP   SEQUENCE OF 1047-1167 FROM N.A.
RC   TISSUE=Aortic smooth muscle;
RA   Adams L.D., Werny I., Schwartz S.M.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PLAYS A ROLE IN TIGHT JUNCTIONS AND ADHERENS JUNCTIONS.
CC   -!- SUBUNIT: INTERACTS WITH OCCLUDIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE (BY
CC   SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A1 (SHOWN HERE), C1, A2 AND C2;
CC   ARE PRODUCED BY ALTERNATIVE SPLICING AND ALTERNATIVE INITIATION.
CC   -!- TISSUE SPECIFICITY: THIS PROTEIN IS FOUND IN EPITHELIAL CELL
CC   JUNCTIONS. ISOFORM A1 IS ABUNDANT IN THE HEART AND BRAIN WHEREAS
CC   ISOFORM C1 IS EXPRESSED AT HIGH LEVEL IN THE KIDNEY, PANCREAS,
CC   HEART AND PLACENTA. IN BRAIN AND SKELETAL MUSCLE, ONLY ISOFORM A1
CC   IS DETECTABLE. ISOFORM C1 IS FOUND IN NORMAL AS WELL AS IN MOST
CC   NEOPLASTIC TISSUES WHILE ISOFORM A1 IS PRESENT ALMOST EXCLUSIVELY
CC   IN NORMAL TISSUE.
CC   -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITION 1086
CC   ONWARD DUE TO A FRAMESHIFT.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; L27476; AAA61300.1; ALT_FRAME.
DR   EMBL; AF177533; AAD20387.2; -.
DR   EMBL; AF043195; AAD20387.2; JOINED.
DR   EMBL; AF043196; AAD20387.2; JOINED.
DR   EMBL; AF043197; AAD20387.2; JOINED.
DR   EMBL; AF177518; AAD20387.2; JOINED.
DR   EMBL; AF177519; AAD20387.2; JOINED.
DR   EMBL; AF177520; AAD20387.2; JOINED.
DR   EMBL; AF177521; AAD20387.2; JOINED.
DR   EMBL; AF177522; AAD20387.2; JOINED.
DR   EMBL; AF177523; AAD20387.2; JOINED.
DR   EMBL; AF177524; AAD20387.2; JOINED.
DR   EMBL; AF177525; AAD20387.2; JOINED.
DR   EMBL; AF177526; AAD20387.2; JOINED.
DR   EMBL; AF177527; AAD20387.2; JOINED.
DR   EMBL; AF177528; AAD20387.2; JOINED.
DR   EMBL; AF177529; AAD20387.2; JOINED.
DR   EMBL; AF177530; AAD20387.2; JOINED.
DR   EMBL; AF177531; AAD20387.2; JOINED.
DR   EMBL; AF177532; AAD20387.2; JOINED.
DR   EMBL; AF177533; AAC02527.2; -.
DR   EMBL; AF043196; AAC02527.2; JOINED.
DR   EMBL; AF043197; AAC02527.2; JOINED.
DR   EMBL; AF177518; AAC02527.2; JOINED.
DR   EMBL; AF177519; AAC02527.2; JOINED.
DR   EMBL; AF177520; AAC02527.2; JOINED.
DR   EMBL; AF177521; AAC02527.2; JOINED.
DR   EMBL; AF177522; AAC02527.2; JOINED.
DR   EMBL; AF177523; AAC02527.2; JOINED.
DR   EMBL; AF177524; AAC02527.2; JOINED.
DR   EMBL; AF177525; AAC02527.2; JOINED.
DR   EMBL; AF177526; AAC02527.2; JOINED.
DR   EMBL; AF177527; AAC02527.2; JOINED.
DR   EMBL; AF177528; AAC02527.2; JOINED.
DR   EMBL; AF177529; AAC02527.2; JOINED.
DR   EMBL; AF177530; AAC02527.2; JOINED.
DR   EMBL; AF177531; AAC02527.2; JOINED.

```
DR   EMBL; AF177532; AAC02527.2; JOINED.
DR   EMBL; AF177533; AAD56218.2; -.
DR   EMBL; AF043195; AAD56218.2; JOINED.
DR   EMBL; AF043196; AAD56218.2; JOINED.
DR   EMBL; AF043197; AAD56218.2; JOINED.
DR   EMBL; AF175518; AAD56218.2; JOINED.
DR   EMBL; AF177519; AAD56218.2; JOINED.
DR   EMBL; AF177520; AAD56218.2; JOINED.
DR   EMBL; AF177521; AAD56218.2; JOINED.
DR   EMBL; AF177522; AAD56218.2; JOINED.
DR   EMBL; AF177523; AAD56218.2; JOINED.
DR   EMBL; AF177524; AAD56218.2; JOINED.
DR   EMBL; AF177525; AAD56218.2; JOINED.
DR   EMBL; AF177526; AAD56218.2; JOINED.
DR   EMBL; AF177527; AAD56218.2; JOINED.
DR   EMBL; AF177528; AAD56218.2; JOINED.
DR   EMBL; AF177529; AAD56218.2; JOINED.
DR   EMBL; AF177532; AAD56218.2; JOINED.
DR   EMBL; AF177533; AAD56218.2; -.
DR   EMBL; AF043196; AAD56219.2; JOINED.
DR   EMBL; AF043197; AAD56219.2; JOINED.
DR   EMBL; AF177518; AAD56219.2; JOINED.
DR   EMBL; AF177519; AAD56219.2; JOINED.
DR   EMBL; AF177520; AAD56219.2; JOINED.
DR   EMBL; AF177521; AAD56219.2; JOINED.
DR   EMBL; AF177522; AAD56219.2; JOINED.
DR   EMBL; AF177523; AAD56219.2; JOINED.
DR   EMBL; AF177524; AAD56219.2; JOINED.
DR   EMBL; AF177525; AAD56219.2; JOINED.
DR   EMBL; AF177526; AAD56219.2; JOINED.
DR   EMBL; AF177527; AAD56219.2; JOINED.
DR   EMBL; AF177528; AAD56219.2; JOINED.
DR   EMBL; AF177529; AAD56219.2; JOINED.
DR   EMBL; AF177532; AAD56219.2; JOINED.
DR   EMBL; AF177533; AAD56219.2; -.
DR   EMBL; AF083892; AAC33121.1; -.
DR   EMBL; AF083893; AAC33122.1; -.
DR   EMBL; U84581; AAB41794.1; -.
DR   HSSP; P31016; 1BE9.
DR   InterPro; IPR000619; Guanylate_kin.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00625; Guanylate_kin; 1.
DR   Pfam; PF00595; PDZ; 3.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 3.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS008856; GUANYLATE_KINASE_1; FALSE_NEG.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 3.
DR   PROSITE; PS50002; SH3; 1.
KW   Tight junction; SH3 domain; Repeat; Membrane; Alternative splicing;
KW   Alternative initiation.
FT   CHAIN         1   1190       TIGHT JUNCTION PROTEIN ZO-2, ISOFORMS
FT                                A1/A2.
FT   CHAIN        24   1190       TIGHT JUNCTION PROTEIN ZO-2, ISOFORMS
FT                                C1/C2.
FT   INIT_MET     24     24       FOR ISOFORMS C1/C2.
FT   DOMAIN       33    120       PDZ 1.
FT   DOMAIN      307    385       PDZ 2.
FT   DOMAIN      509    590       PDZ 3.
FT   DOMAIN      604    669       SH3.
FT   DOMAIN      690    876       GUANYLATE KINASE.
FT   DOMAIN     1162   1165       POLY-GLU.
FT   VARSPLIC    961   1108       MISSING (IN ISOFORM A2/C2).
FT   CONFLICT    411    411       N -> T (IN REF. 1).
FT   CONFLICT    782    782       I -> V (IN REF. 1).
FT   CONFLICT    808    808       P -> S (IN REF. 1).
FT   CONFLICT    812    814       FFN -> SFT (IN REF. 1).
FT   CONFLICT    822    822       K -> N (IN REF. 1).
FT   CONFLICT    822    822       N -> D (IN REF. 1).
FT   CONFLICT    829    829       K -> N (IN REF. 1).
FT   CONFLICT    834    834       N -> D (IN REF. 1).
FT   CONFLICT    842    842       Q -> H (IN REF. 1).
FT   CONFLICT    996    996       P -> S (IN REF. 1).
FT   CONFLICT   1092   1095       ARLQ -> GPGL (IN REF. 1).
FT   CONFLICT   1136   1136       S -> N (IN REF. 5).
FT   CONFLICT   1155   1158       GSYG -> RSFC (IN REF. 5).
FT   CONFLICT   1165   1167       EYR -> IRS (IN REF. 5).
SQ   SEQUENCE   1190 AA;  133971 MW;  BE2BE6F1B1467058 CRC64;
```

```
Query Match             4.9%;  Score 254.5;  DB 1;  Length 1190;
Best Local Similarity  21.5%;  Pred. No. 1.5e-05;
Matches  213;  Conservative  144;  Mismatches  394;  Indels  239;  Gaps  50;

Qy       55 EEVLHSPRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNP-DVYTLVTGLQPPDVDF 112
              : |  |||                                 :  | :  ::   |
Db       94 EDVLHS-FAVQQLRKSGKVAAIVVKRPRKVQVAALQA----SPPLDQDDRAFEVMDEFDG 148

Qy      113 SNF-SGLMETSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLH 171
             |||    | |  |  :|    ||||   ||::|   :| || |: |
Db      149 RSFRSGYSERSRLNS-HGGRSRSWED--SPERGRPH---ERARSRERDLSRDRSRGRSLE 202

Qy      172 Q-LEADHSRMKREVSAHFHEVLRLKDEMLSLSL-HYSNALQEKELAASRCRSLQEELYLL 229
             : |:||:                    |  ::: :|  :   | |
Db      203 RGLDQDHARTRD-------RSRGRSLERGLDHDFGPSRDRDRDRSRGRSID------ 246

Qy      230 KQELQRANMVSSCELELQEQSLRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQ 289
                |  ||               :                        |
Db      247 -QDYERAY---------------HRAYDPDYER---------------------AY 265

Qy      290 SLDEARGSRQELVERIHSLRERAVAAERQRQYWEEKEQTLLQFQKSKMACQLYREKVNA 349
             | | ||:                  |   |    |  || |
Db      266 SPEYRRGAR-------HDARSRG-PRSRSREBPHSRSPSPEPRGRPGIGVLLMKSRAN- 316

Qy      350 LQAAQVCELQKERDQAYSARDSAQ---REISQS-LVEKDSLRRQ---VFELTDQVCE--LR 400
                     :   :|:  |||:|     | :|   | :  | |||        | |  |
Db      317 ----------EEYGLRLGSQIFVKEMTRTGLATKDGNLHEGDIILKINGTVTENMSL 363

Qy      401 TQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDDSDCSL--VSSTES---- 453
             | |:|  |  |     |     ||:|         :|  |
Db      364 TDARKLIEKSRGKL-QLVVLRD---SQQTLINIPSL--NDSDSEIEDISEIESNRSFS 415

Qy      454 ----QLLSDLSATSSRELVDSFRSSPAPSSQQSLYKRVAEDFGEPWSFSSCLEIPEG 508
                  || |  |   | |              :  ||  : :|    ||
Db      416 PEERRHQYSDYDHSSSEEKLKERPSSREDTPSRLS-----RMGATPTPFKS------ 461

Qy      509 DPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRRPARRIL- 566
              | |     |  |   | | :  |  ||:|::  || | |    |     |   ||
Db      462 -TGDIAGTVVPETNKEPRYQEPPAP------QPKAAPR------TFLRFSPEDEAIYG 507

Qy      567 SQVTMLAF-QGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEP 625
              | :|| :  : |||      ||  |||| ||: ||      :|    | :      ::
Db      508 PNTKMVRFKKGDSV--GLRLAGGNDVGIFVAGIQEGTSAEQEGLQEGDQILKVNTQDFRG 565

Qy      626 LFKAVLEDTTLEEAVGLLRRVDG-FCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVN 684
             | | |: |                 |   :       | |: |  |      |  | |:
Db      566 L---VREDAVLY---LLEIPKGEMVTILAQSRADVYRDIL------ACGRGDSFFIRSH 612

Qy      685 LAMEGRAKGELQVHCNEVLHVDTMFQG-CGCWAHRVNSYTMKDTAAHGTIPNYSRAQQ 743
              |  ||  | |        ||||  |  |||  ||| :   |  |  |||| ::||  |
Db      613 FECEKETPQSLAFTRGEVFRVVDTLYDGKLGNWLAVRIGNELEK----GLIPNKSRAEQ 667

Qy      744 QLIALIQDMTQQCTVIRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSST 803
              :  ::                  | |    || :| |  |: |  |||        |
Db      668 --MASVQN------AQRDNAGDRADFWRM-----RGQRSGVKKNLRKSREDLTAVVSVST 714

Qy      804 CFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKICLLQGFKKCLAEYLSQEE 863
              |               : ||              |||  |  ||    ||
Db      715 KFPA-------YERVLLREAGFKRPVVLF-GPIADIAMEKLA---------------NE 750

Qy      864 YEAWSORGDIIQEGEVSGGRCWVTR-HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIV 922
               |  :|   | ||   |   | :|  | | :  :|:     |: |  :|  : ||| ||
Db      751 LPDWFQTAKTEPKDAGSEKSTGVVRLNTVRQIIEQDKHALLDVTPKAVDLLNYTQWFPIV 810

Qy      923 IHVSVNEKMAKKLKKGLQRL-----GTSEEQLLEAARQEEGLDRAPCLY----SSLA 971
             |   :  |                    |::|    |::|  |  |          ||
```

```
Db       811  IFFNPDSRQGVKTMR--QRLNPTSNKSSRKLFDQANKL----KKTCAHLFTATINLNSA 863
                 |  |:  :. | |||      :| |       :  | :| |   :|  :|
Qy       972  PDGWSDLDGLLSCVRQAIADEQKKVVWTEQ 1001
                 |    |  ::: || :| ||  :    :| ||
Db       864  NDSW------FGSLKDTIQHQQGEAVWVSE 887


RESULT  13
PLE1_HUMAN
ID   PLE1_HUMAN    STANDARD;     PRT;   4684 AA.
AC   Q15149; Q16640; Q15148;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1).
GN   PLEC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Placenta;
RX   MEDLINE=96210632; PubMed=8633055;
RA   Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT   "Human plectin: organization of the gene, sequence analysis, and
RT   chromosome localization (8q24).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.
RX   MEDLINE=96312447; PubMed=8698233;
RA   McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,
RA   Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owaribe K.,
RA   McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,
RA   Uitto J.;
RT   "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:
RT   cDNA cloning and genomic organization.";
RL   Genes Dev. 10:1724-1735(1996).
RN   [3]
RP   VARIANT MD-EBS 1003-GLN--ALA-1005 DEL.
RX   MEDLINE=97049959; PubMed=8894687;
RA   Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,
RA   Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.;
RT   "Homozygous deletion mutations in the plectin gene (PLEC1) in patients
RT   with epidermolysis bullosa simplex associated with late-onset
RT   muscular dystrophy.";
RL   Hum. Mol. Genet. 5:1539-1546(1996).
RN   [4]
RP   VARIANT MD-EBS LEU-429 INS.
RX   MEDLINE=21090821; PubMed=11159198;
RA   Bauer J.W., Rouan F., Kofler B., Rezniczek G.A., Kornacker I.,
RA   Muss W., Hametner R., Klausegger A., Huber A., Pohla-Gubo G.,
RA   Wiche G., Uitto J., Hintner H.;
RT   "A compound heterozygous one amino-acid insertion/nonsense mutation in
RT   the plectin gene causes epidermolysis bullosa simplex with plectin
RT   deficiency.";
RL   Am. J. Pathol. 158:617-625(2001)
CC   -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC       MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC       HEMIDESMOSOMES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO
CC       MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE
CC       CROSSLINKING AND STABILIZATION OF CYTOSKELETAL INTERMEDIATE
CC       FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC   -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN
CC       MUSCLE, HEART, PLACENTA AND SPINAL CORD.
CC   -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC       VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
CC       AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC   -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC       INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC   -!- DISEASE: DEFECTS IN PLEC1 ARE THE CAUSE OF EPIDERMOLYSIS BULLOSA
CC       WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-MD), AN AUTOSOMAL RECESSIVE
CC       DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL
CC       OF THE HEMIDESMOSOME AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.
CC   -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 33 PLECTIN REPEATS.
CC   -!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z54367; CAA91196.1; -.
DR   EMBL; U53204; AAB05427.1; -.
DR   EMBL; U63610; AAB05428.1; -.
DR   EMBL; U63609; AAB05428.1; JOINED.
DR   EMBL; X97053; CAA65765.1; -.
DR   HSSP; Q01082; 1BKR.
DR   MIM; 601282; -.
DR   MIM; 226670; -.
DR   InterPro; IPR001589; Actinin_act_bind.
DR   InterPro; IPR001715; Calponin_hom.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00681; Plectin_repeat; 19.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00250; PLEC; 33.
DR   SMART; SM00150; SPEC; 5.
DR   PROSITE; PS00019; ACTININ_1; FALSE_NEG.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
KW   Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW   Phosphorylation; Alternative splicing; Epidermolysis bullosa;
KW   Disease mutation.
FT   DOMAIN       1   1470
FT   DOMAIN    1471   2755       CENTRAL FIBROUS ROD DOMAIN.
FT   DOMAIN    2756   4684
FT   DOMAIN     175    400       ACTIN-BINDING.
FT   DOMAIN     179    282       CH 1.
FT   DOMAIN     295    397       CH 2.
FT   REPEAT     645    710       SPECTRIN 1.
FT   REPEAT     740    824       SPECTRIN 2.
FT   REPEAT     837    930       SPECTRIN 3.
FT   REPEAT    1315   1415       SPECTRIN 4.
FT   DOMAIN    1469   2756       COILED COIL (POTENTIAL).
FT   REPEAT    2826   2863       PLECTIN 1.
FT   REPEAT    2864   2901       PLECTIN 2.
FT   REPEAT    2902   2939       PLECTIN 3.
FT   REPEAT    2940   2977       PLECTIN 4.
FT   REPEAT    2981   3015       PLECTIN 5.
FT   REPEAT    3116   3153       PLECTIN 6.
FT   REPEAT    3154   3191       PLECTIN 7.
FT   REPEAT    3192   3229       PLECTIN 8.
FT   REPEAT    3230   3267       PLECTIN 9.
FT   REPEAT    3268   3305       PLECTIN 10.
FT   REPEAT    3306   3343       PLECTIN 11.
FT   REPEAT    3485   3522       PLECTIN 12.
FT   REPEAT    3523   3560       PLECTIN 13.
FT   REPEAT    3561   3598       PLECTIN 14.
FT   REPEAT    3599   3636       PLECTIN 15.
FT   REPEAT    3640   3674       PLECTIN 16.
FT   REPEAT    3820   3857       PLECTIN 17.
FT   REPEAT    3858   3895       PLECTIN 18.
FT   REPEAT    3896   3933       PLECTIN 19.
FT   REPEAT    3934   3971       PLECTIN 20.
```

```
Qy  195 KDEMLSLSLHYSNALQEKELAASRCRSLQEEYLLKQEL-QRANMVSSCELELQEQ---  249
            |   |    :  |||| ::||| |:   :  ||: ||
Db 2463 KAEAELLQ---------QQKELAQEQAARRLQEDKEQMAQQLAEETQGFQRTLAERQRQLEM 2515

Qy  250 ------SLRTAS--DQESGDEELNRLKEE---NEKLRSLTFSLAEKDILEQSLDEAR  295
Db 2516 SAEAERIKLRVAEMSRAQARAEEDAQRFRKQAEEIGEKLHRTELATQEKVTLVQTLEIQR 2575

Qy  296 GSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMAC---QLYREKVNALQ  351
Db 2576 QQSDHDAER---LREAIAELEREKEKL-QQEAKLLQLKSEEMQTVQQEQLLQETQALQQ 2630

Qy  352 AQVCE-----LQKER------                                       385
Db 2631 SFLSEKDSLLQRERFIEQEKAKLEQLFQDEVAKAQQLREEQORQQQMEQEBQRLVASME 2690

Qy  386 --RRQVFELTDQVCELRTQLRLQLQAEPPGVLKQEARTREPCCPREKQRLVRMHAICPRDDS 443
Db 2691 EARRRQHEAEEGVRRKQEELQQLE------QQRRQQEELLAEENQRLREQLQLLEEQH   2742

Qy  444 DCSLVSS--TESQLLSDLSATSSRELVDSFRSSSPAPPSQQSLYRKVAEDFGEEPWSFS  500
Db 2743 RAALAHSEEVTASQVAATKTLPNGRDALDGPAAEAEPEHSFDGLRRKVSAQRLQEAGILS 2802

Qy  501 SCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRR 560
Db 2803 A----------------EELQRLAQGHTTVDE---LARRE                    2823

Qy  561 PARRLSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADOMALRPGTQIVMVDY  620
Db 2824 DVRHYL------QGR-----SSIAGLLLKATNEKLSVYAALQRQLLSPGTALILLEA   2869

Qy  621 EAS------EPLFKAVLEDTLEEAV-GLLRRVDGFCCLSVVNTDGYK---------    662
Db 2870 QAASGFLLDPVRNRRL--TVNEAVKEGVVGPELHHKLLSAERAVTGYKDPYTGQQISLF 2926

Qy  663 RLLQD------DSFYIRVNLAMEGRAKGELQVHCNEVLHVT                   706
Db 2927 QAMQKGLIVREHGIRLLEAQIATGGVIDPVHSRVPVDVAYR---RGYFDEEMNRVLADP 2983

Qy  707 DTMFQCGCGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALLQD-MTQQC--TVTRKPS 763
Db 2984 ----QLLERCVEDPETGLCLLPLTDKAA                                3026

Qy  764 SGGPQKLVRIVSMDKAKASPLRLSFDRGQLDP--SRMEGSSTCFWAESCLTLVPYTLVWP 821
Db 3027 KGG------ELVYTD---SEARDVFKATVSAPFGKFQGKTVTIW--EIINSEYFT---  3071

Qy  822 HRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEYEAWSQRGDIIQEG     877
Db 3072 ---AEQRDLLRQFRTGRITVEKII-----KIIITVVEEQE-----QKGRLCFEG     3113
```

```
RESULT 14
ID   ZO3_MOUSE     STANDARD;     PRT;    905 AA.
AC   Q9QXY1;
     Q9QXY1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE   occludens 3 protein) (Tight junction protein 3).
GN   TJP3 OR ZO3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=20069797; PubMed=10601346;
RA   Itoh M., Furuse M., Morita K., Kubota K., Saitou M., Tsukita S.;
RT   "Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2,
```

```
FT   REPEAT    3975  4008  PLECTIN 21.
FT   REPEAT    4063  4100  PLECTIN 22.
FT   REPEAT    4101  4138  PLECTIN 23.
FT   REPEAT    4139  4176  PLECTIN 24.
FT   REPEAT    4177  4214  PLECTIN 25.
FT   REPEAT    4218  4252  PLECTIN 26.
FT   REPEAT    4265  4305  PLECTIN 27.
FT   REPEAT    4319  4356  PLECTIN 28.
FT   REPEAT    4408  4445  PLECTIN 29.
FT   REPEAT    4446  4483  PLECTIN 30.
FT   REPEAT    4484  4521  PLECTIN 31.
FT   REPEAT    4522  4559  PLECTIN 32.
FT   REPEAT    4560  4597  PLECTIN 33.
FT   DOMAIN    4250  4300  BINDING TO INTERMEDIATE FILAMENTS (BY
FT                         SIMILARITY).
FT   DOMAIN    4625  4640  4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT   MOD_RES   4539  4539  PHOSPHORYLATION (BY CDC2) (BY
FT                         SIMILARITY).
FT   VARSPLIC     1   174  MVAGMLMPRDQLRAIYEVLFREGVMVAKKDRRPRSLHPHVP
FT                         GVTNLQVMRAMASLRARGLVRETFAWCHFFWYLTNEGIAHL
FT                         RQYLHLPPEIVAASLQRVRRPVAMVMPARRTPHVQAVQGPL
FT                         GSPPKRGPLPTEEQRLYRRKELEEVSPETPVVPATTQRTLA
FT                         RPGPEPAPAT -> MSGEDAEVRAVSEDVSNGSSGSPSPGD
FT                         TLPWNLGKTQRSRSGGAGSNGSVLDPAERAVIRIA (IN
FT                         ISOFORM 2 AND ISOFORM 3).
FT   VARSPLIC   409   412  MISSING (IN ISOFORM 3).
FT   VARIANT    429   429  L -> LL (IN MD-EBS).
FT                         /FTId=VAR_011336.
FT   VARIANT   1003  1005  MISSING (IN MD-EBS).
FT                         /FTId=VAR_011337.
FT   CONFLICT   185   185  S -> F (IN REF. 2).
FT   CONFLICT   259   259  N -> D (IN REF. 2).
FT   CONFLICT   550   550  H -> N (IN REF. 2).
FT   CONFLICT   560   560  I -> V (IN REF. 2).
FT   CONFLICT   706   706  Q -> R (IN REF. 2).
FT   CONFLICT   886   886  N -> Y (IN REF. 2).
FT   CONFLICT  1002  1002  V -> A (IN REF. 2).
FT   CONFLICT  1309  1309  L -> L (IN REF. 2).
FT   CONFLICT  1321  1321  L -> L (IN REF. 2).
FT   CONFLICT  1334  1334  V -> L (IN REF. 2).
FT   CONFLICT  1534  1534  I -> M (IN REF. 2).
FT   CONFLICT  1662  1662  A -> T (IN REF. 2).
FT   CONFLICT  1688  1690  WLC -> RLR (IN REF. 2).
FT   CONFLICT  1767  1767  Q -> E (IN REF. 2).
FT   CONFLICT  1789  1789  L -> R (IN REF. 2).
FT   CONFLICT  1910  1910  K -> R (IN REF. 2).
FT   CONFLICT  2154  2154  S -> N (IN REF. 2).
FT   CONFLICT  2160  2160  R -> N (IN REF. 2).
FT   CONFLICT  2215  2215  R -> Q (IN REF. 2).
FT   CONFLICT  2244  2244  S -> A (IN REF. 2).
FT   CONFLICT  3027  3027  K -> E (IN REF. 2).
FT   CONFLICT  3310  3310  E -> A (IN REF. 2).
FT   CONFLICT  3361  3361  F -> L (IN REF. 2).
FT   CONFLICT  3408  3408  F -> L (IN REF. 2).
FT   CONFLICT  3447  3447  S -> A (IN REF. 2).
FT   CONFLICT  3531  3531  G -> A (IN REF. 2).
FT   CONFLICT  3580  3580  R -> S (IN REF. 2).
FT   CONFLICT  3589  3589  K -> Q (IN REF. 2).
FT   CONFLICT  3596  3596  E -> H (IN REF. 2).
FT   CONFLICT  3616  3616  N -> H (IN REF. 2).
FT   CONFLICT  3686  3686  V -> A (IN REF. 2).
FT   CONFLICT  3786  3786  G -> A (IN REF. 2).
FT   CONFLICT  3808  3808  K -> E (IN REF. 2).
FT   CONFLICT  3816  3816  G -> A (IN REF. 2).
```

```
Query Match      4.8%;    Score 249;    DB 1;    Length 4684;
Best Local Similarity   22.4%;    Pred. No. 0.00015;
Matches 187;   Conservative 112;   Mismatches 319;   Indels 218;   Gaps 34;

Qy  135 LQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRL 194
            ||||     :  ||  |  :  :  |  | ||| : : | | ||
Db 2403 LQEEAEKMKQVAQEAARLRQLAEEDLAQQRALAEKMLKEKMQAVQEATRL           2462
```

```
RT       and ZO-3, with the COOH termini of claudins.";
RL       J. Cell Biol. 147:1351-1363(1999).
CC   -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.
CC   -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC
DR   EMBL; AF157006; AAF24175.1; -.
DR   HSSP; P31016; 1BFE.
DR   MGD; MGI:1351650; Tjp3.
DR   InterPro; IPR000619; Guanylate_kin.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00625; Guanylate_kin; 1.
DR   Pfam; PF00595; PDZ; 3.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 3.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50002; SH3; FALSE_NEG.
DR   PROSITE; PS50106; PDZ; 3.
KW   Tight junction; SH3 domain; Repeat; Membrane.
FT   DOMAIN        11      93       PDZ 1.
FT   DOMAIN       187     264       PDZ 2.
FT   DOMAIN       368     434       PDZ 3.
FT   DOMAIN       467     540       SH3.
FT   DOMAIN       654     754       GUANYLATE KINASE.
SQ   SEQUENCE   905 AA;  99324 MW;  B787BA159266IFEE CRC64;
```

```
Query Match             4.8%;  Score  246.5;  DB 1;  Length 905;
Best Local Similarity  24.2%;  Pred. No. 2.7e-05;
Matches  142;  Conservative  91;  Mismatches  225;  Indels  129;  Gaps  25;

Qy       448  VSSTESQLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRKVAEDFGEEPWSFSSCLEIPE  507
              || :||:|| :|: ||:| | ||||          |
Db       275  VSDSDSSLMEDISDLTS-EL-------SQAPPS--------------------HVP-    302

Qy       508  GDPGALPGAKAGDPHLDYELLDTADLPQLESSL------QPVSPGRLDVSESGVLMRRRP   561
              :|      |||             :|      |:|:|                  :|
Db       303  --PPPLKGQRSPEDSQTDSPVETPQPRRRERSVNSRAIAEPESPG----ESRYDIYRVP   355

Qy       562  ARRILS------QVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQI  615
              :|:     |:|| ||||| ||| ||| |||:| |||:|||:|| :||:
Db       356  SRQSLEDRGYSPDTRVVSFPKGASI-GLRLAGGNDVGIFVSGVQAGSPADGQGIQEGDEI  414

Qy       616  VMVDYEASEPLFKAVLEDTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATS   675
              :|:   || |  ||| | ||| |:|| |  :|||    | : | || ||||:
Db       415  LQVN------GMPFRNLTREEAVQFLLGLPPGEDMELVTQSKTGHSLRRWSQSRV---   463

Qy       676  GDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQGCGCCWHAHRVNSYTMKDTAA----  731
              ||||||| |||||||:||  ||  |  |:| ||||| ||:|  |||||:|| || ||
Db       464  GDSFYIRTHFELEPSPPYGLGFTRGDVFHVDTLYPGSGPGHGH--SSHGGLWLLAARMGR  521

Qy       732  -----HGTIPNYSRAQQLIALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAKASPLR  785
              |||||||:|| ||:|||||:|:|| ||    ||:  | |||:::|||::|
Db       522  DLREQERGVIPNQSRAEQ-LASLE------AAQRAAGVPPG-ASVGSNPRAEFWRLR  570

Qy       786  LSFDRGQLDPSRMEGSSTCFWAESCLT----LVPYTLVWPHRPARPRPVLLVPRAVGKIL  841
              | :| |||           ||| |:  ||||:|::: ||||||:|
Db       571  -SLRRGTKKASTQRSREDL----SALTRQGHYPPYPYERVLREASFKRPVVIL-GPVADIA  624

Qy       842  SEKLCLLQGFKKCLAEYLSQEEYFAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHA  901
```

```
Db       625  MKKLTEMPEFEIAESMSRIDSPSKIIKLD-----------TVRVIAERDKHA  667
              :||     ||: |: :  |::                   :|:|| :|| ||

Qy       902  LLDVQLDSVCTLHRMDIFPIVHVSVNEKMA-KKLKKGLQRLGTSEEQLLEAARQEEGDL  960
              ||||          |:    : :|||||   ||||:|||:|||||||||||||||::||

Db       668  LLDVTPSAIERLNYVQYYPIVIFCAPESRPALKALREWLAPASRSSRRLYAQAQK---L  724

Qy       961  DR-APCLYSSLAP-------DGWSDLDGLLSCVRQAIADEQKKVVWTEQ  1001
              || || |  ||        |||| |    ||:|| |:   :|:::|

Db       725  QKHSGHLFTATIPLHGTSDSW------YQEVKAVIQQQARPIWTAE  765
```

```
RESULT  15
MYH9_HUMAN
ID   MYH9_HUMAN     STANDARD;      PRT;   1960 AA.
AC   P35579; Q60805;
DT   01-JUN-1994 (Rel. 29, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE   type A) (Nonmuscle myosin heavy chain-A) (NMMHC-A).
GN   MYH9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
RA   Bagguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA   Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA   Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA   Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA   Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
RA   Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA   Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA   Grafham D., Griffiths M.N., Hall R., Hall R., Hall-Tamlyn G.,
RA   Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA   Kimberley R., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA   Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA   Matthews L., Mccann O.T., Mcclay J., Mclaren S., Mcmurray A.A.,
RA   Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA   Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA   Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA   Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA   Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA   Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA   Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA   Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA   Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA   Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA   Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA   Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,
RA   Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA   Sloan D., Song L., Wang J., Wang Y., Wang Y., White J., Willingham D.,
RA   Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA   Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA   Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA   Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA   Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,
RA   Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA   Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA   Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA   Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA   Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA   O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA   Khan A.S., Lane L., Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
RN   [2]
RP   SEQUENCE OF 1-1337 FROM N.A.
```

```
RX   MEDLINE=92003925; PubMed=1912569;
RA   Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
RA   Arnaout M.A., Clayton L.K., Tenen D.G.;
RT   "Cellular myosin heavy chain in human leukocytes: isolation of 5'
RT   cDNA clones, characterization of the protein, chromosomal
RT   localization, and upregulation during myeloid differentiation.";
RL   Blood 78:1826-1833(1991).
RN   [3]
RP   SEQUENCE OF 1-715 FROM N.A.
RX   MEDLINE=91316803; PubMed=1860190;
RA   Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA   Gdula D., Adelstein R.S., Weir L.;
RT   "Human nonmuscle myosin heavy chains are encoded by two genes located
RT   on different chromosomes.";
RL   Circ. Res. 69:530-539(1991).
RN   [4]
RP   SEQUENCE OF 714-1960 FROM N.A.
RX   MEDLINE=90198958; PubMed=1967836;
RA   Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;
RT   "Human nonmuscle myosin heavy chain mRNA: generation of diversity
RT   through alternative polyadenylylation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
RN   [5]
RP   VARIANT DFNA17 HIS-705.
RX   MEDLINE=20489856; PubMed=11023810;
RA   Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
RA   Mhatre A.N.;
RT   "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
RT   nonmuscle myosin MYH9.";
RL   Am. J. Hum. Genet. 67:1121-1128(2000).
RN   [6]
RP   VARIANTS MHA/FTNS/SBS K-93; C-702; C-1165; H-1424 AND K-1841.
RX   MEDLINE=20428192; PubMed=10973259;
RA   Seri M., Cusano R., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
RA   Ghiggeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
RA   Iolascon A., Zelante L.L., Savoia A., Balduini C.L., Noris P.,
RA   Magrini U., Belletti S., Heath K.E., Babcock M., Glucksman M.J.,
RA   Aliprandis E., Bizzaro N., Desnick R.J., Martignetti J.A.;
RT   "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
RT   Sebastian syndromes.";
RL   Nat. Genet. 26:103-105(2000).
RN   [7]
RP   VARIANTS MHA ILE-1155 AND LYS-1841.
RX   MEDLINE=20428193; PubMed=10973260;
RA   Kelley M.J., Jawien W., Ortel T.L., Korczak J.F.;
RT   "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
RT   May-Hegglin anomaly.";
RL   Nat. Genet. 26:106-108(2000).
CC   -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC       CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC       CAPPING.
CC   -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC       CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC       REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY
CC       (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
CC       BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
CC   -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME
CC       (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
CC       BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
CC       WITH ADDITIONAL ALPORT-LIKE CLINICAL FEATURES OF SENSORINEURAL
CC       DEAFNESS, CATARACTS AND NEPHRITIS.
CC   -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME
CC       (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
CC       BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
CC   -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS
CC       CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND
CC       COCHLEOSACCULAR DEGENERATION.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z82215; CAB05105.1; -.
DR   EMBL; M81105; AAA59888.1; -.
DR   EMBL; M69180; AAA61765.1; -.
DR   EMBL; M31013; AAA36349.1; -.
DR   HSSP; P08799; 1LVK.
DR   MIM; 160775; -.
DR   MIM; 153640; -.
DR   MIM; 155100; -.
DR   MIM; 603622; -.
DR   MIM; 605249; -.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Alkylation; Multigene family; Disease mutation;
KW   Deafness.
FT   DOMAIN        1    778       MYOSIN HEAD-LIKE.
FT   DOMAIN      779    808       IQ.
FT   DOMAIN      837   1926       COILED COIL (POTENTIAL).
FT   NP_BIND     174    181       ATP (POTENTIAL).
FT   DOMAIN      654    676       ACTIN-BINDING.
FT   MOD_RES     694    694       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES     704    704       ALKYLATION (SH-2) (POTENTIAL).
FT   VARIANT      93     93       N -> K (IN MHA).
FT                                /FTId=VAR_010791
FT   VARIANT     702    702       R -> C (IN FTNS).
FT                                /FTId=VAR_010792
FT   VARIANT     705    705       R -> H (IN DFNA17).
FT                                /FTId=VAR_010793
FT   VARIANT    1155   1155       T -> I (IN MHA).
FT                                /FTId=VAR_010794
FT   VARIANT    1165   1165       R -> C (IN SBS).
FT                                /FTId=VAR_010795
FT   VARIANT    1424   1424       D -> H (IN FTNS).
FT                                /FTId=VAR_010796
FT   VARIANT    1841   1841       E -> K (IN MHA).
FT                                /FTId=VAR_010797
FT   CONFLICT     53     55       EAI -> RGH (IN REF. 3).
FT   CONFLICT    660    660       T'-> S (IN REF. 3).
FT   CONFLICT    869    869       T -> M (IN REF. 4).
FT   CONFLICT    931    931       C -> Y (IN REF. 4).
FT   CONFLICT   1240   1241       KG -> GR (IN REF. 4).
FT   CONFLICT   1350   1350       E -> EE (IN REF. 2).
FT   CONFLICT   1764   1764       T -> A (IN REF. 2).
FT   CONFLICT   1771   1771       S -> G (IN REF. 2).
SQ   SEQUENCE   1960 AA;  226531 MW;  588F84BB8C106E6F CRC64;

Query Match             4.7%;  Score 240.5;  DB 1;  Length 1960;
Best Local Similarity   21.1%;  Pred. No. 0.00014;
Matches  238;  Conservative  176;  Mismatches  359;  Indels  353;  Gaps  59;
```

```
Qy    14 LDEETLWEMMESHRH---RIVRCICPS------RLTPYLRQAKVLCQLDEEEVLHSPR-                      62
          | :|:|::|:| ||                || |||  ||||  ||||
Db   649 LYKEQLAKLMATLRNTNPNFVRCIIPNHEKKAGKLDPHL----VLDQLRCNGVLEGIRIC                        704

Qy    63 -----------------LTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTL-                          102
                             || :::|            | ||    ||: |   ::||
Db   705 RQGFPNRVVFQEFRQRYEILTPNSIPKGFM------DGKQACVLMIKALEL-DSNLYRIG                        757

Qy   103 --------------VTGLQPDVDFSNFSGLMETSKLTECLAG-----------AIGSLQEELN                     140
                         : |:|| |             |:|:| :|          ||:|
Db   758 QSKVFFRAGVLAHLEEERDL------------KITDVIIGFQACCRGYLARKAFAKRQQQLT                      807

Qy   141 QEKGQKEVLLRRCQQLQEHLGLAE------TRAEGLHQLEADHSRMKREVSAHFHEVLR                        193
          :|  |||          |:       |:|         ||:||:::
Db   808 AMK-----VLQRNC--AAYLKLRNWQWRLFTKVKPLLQV---SRQEEMMAKEEELVK                          856

Qy   194 LKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEQSLR-                        252
          ::::|:     | |           |:| ||   | | :|
Db   857 VREKQLAA----ENRLTEME------TLQSQLMAEKLQLQE-QLQAETELCAEAEELRA                        904

Qy   253 --TASDQESGDEEL-------NRLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQELVERI                      305
          || ||  ||     |  |::      || ||  :|            :|
Db   905 RLTAKKQEL--EEICHDLEARVEEEERCQHLQ---AEKKKMQQNIQELEEEESAR                            959

Qy   306 HSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCE----LQKER                       361
          | |||     |       ||||      ||  :| |          :|
Db   960 QKLQLEKVTTEAKLKKL--EEEQILLEDQN----CKLAKEK-KILEDRIAEFTTNLTEEE                        1012

Qy   362 DQAYS-----------ARDSAQR---EISQSLVEKDS--LRROVFELTDQVC                               397
          :|:             :|      | |  |:       | || |::|   |
Db  1013 EKSKSLAKLKNKHEAMITDLEERLRREEKQRQELEKTRRKLEGDSTDLSDQIAELQAQIA                       1072

Qy   398 ELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLS                        457
          |: ||: | ||:  |:   |   ::                   ||| :|
Db  1073 ELKMQLAKKEEELQAAL--ARVEEEAAQKNMALKIREL----------ESQ-IS                            1114

Qy   458 DLSATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAK                        517
          |:                                            ||
Db  1115 EL-----QEDLESERASRNKAEKQK--------RDLGEE--------LEA-                                1143

Qy   518 AGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGV----------LMRRRPARRI                        565
                                          |:  :|
Db  1144 ------LKTELEDTLDSTAAQQELRSKREQEVNILKKTLEEEAKTHEAQIQEMRQKHSQ--                       1196

Qy   566 LSQVTMLAFQGDALLEQISVIGGNL----------TGIFIHRV------TPGSAADQMAL                       609
          :|                          || | |        ||
Db  1197 -AVEELAEQ----LEQTKRVKANLEKAKQTLENERGELANEVKVLLQGKGDSEHKRKV                         1250

Qy   610 RPGTQIVMVDYEASEPLFKAVLEDTT----LEEAVGLLRRVDGFCCLSVKVNTD--GYK                         662
          | |                  ::           | : |||    |||   | :|
Db  1251 EAQLQELQVKFNEGERVRTELADKVTKLQVELDNVTGLLSQSDS--KSSKLTKDFSALE                        1307

Qy   663 RLLQDLE------------AKVATSGDSFYIRVNLAMEGRAKGELQ-----VHCN                           700
          :|                 :|||          |        |:
Db  1308 SQIQDTQELLQENRQKLSLSTKLKQVEDEKNSF--REQLEEEEAKHNLEKQIATLHA-                         1364

Qy   701 EVLHVTDTMFQGCGCWH-AHRVNSYTMKD----------TAAHGTI-PNYSRAQQQLIA                        747
          || | |              |  :::|             || :: ||  ||:
Db  1365 QVADMKKKMEDSVGCLETAEEVKRKLQKDLEGLSQRHEEKVAAYDKLEKTKTRLQQELDD                        1424

Qy   748 LIQDMTQQ-----CTVTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSST                        803
          |:| ::          :|  | ||             | ||      |:|:
Db  1425 LLVDLDHQRQSACNLEKK-----QKKFDQLLAEEKTISAKYAEERDRAEAEARKETKA-                         1478

Qy   804 CFWAESCLTLVPYTLVPNHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEE                        863
          |                                     :|    |:| |   |
Db  1479 ------------------------------LSLARALEEAMEQKAEL----ERLNKQFRTEM                     1506

Qy   864 YEAWSQRGDI---IQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFP                        920
                       ::|:        ||:    :|  |  |::||
Db  1507 EDLMSSKDDVGKSVHELEKS------KRALEQQVEEMKTQLEELEDELQATEDAK-----                        1555

Qy   921 IVIHVSVN-EKMAKKLKKGLQ-RLGTSEE--QLLEAARQEEGDLD                                       961
```

```
Db  1556 --LRLEVNLQAMKAQFERDLQGRDEQSEEKKQLVRQVREMEAELE 1599
               :.:||  |.: ::|||    |||     ||  | : ||
```

Search completed: June 13, 2002, 09:26:01
Job time: 184 sec

```
                    GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      June 13, 2002, 09:22:22 ; Search time 26.06 Seconds
                                      (without alignments)
                                      3701.984 Million cell updates/sec

Title:       US-09-767-215-2
Perfect score:  5149
Sequence:    1 MGELCRRDSALTALDEETLW.........VRQAIADEQKVVWFEQSPR 1004

Scoring table:  BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters:    283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_71:*
             1: pir1:*
             2: pir2:*
             3: pir3:*
             4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	289	5.6	1745	2	A46431	tight junction-ass
2	282	5.5	1367	2	T13703	tamA protein - fru
3	270.5	5.3	1736	2	A47747	tight junction pro
4	262	5.1	2101	2	A42184	nuclear mitotic ap
5	249	4.8	4574	2	G02520	plectin - human
6	249	4.8	4684	2	A59404	plectin [imported]
7	248.5	4.8	2442	2	T08621	centrosome associa
8	246.5	4.8	1116	2	I54378	gene Xl04 protein
9	245	4.8	746	2	T47237	myosin II heavy ch
10	243.5	4.7	2168	2	T30171	ninein - mouse
11	242.5	4.7	3225	2	I52300	giantin - human
12	242.5	4.7	3259	2	A56539	giantin - human
13	242	4.7	1163	2	JE0366	tight junction pro
14	236.5	4.6	775	1	I46236	tight junction pro
15	235	4.6	1959	1	A33977	myosin heavy chain
16	235	4.6	1961	1	A61231	myosin heavy chain
17	234.5	4.6	4687	2	A39638	plectin - rat
18	230	4.5	1281	2	T00346	hypothetical prote
19	226	4.4	1972	1	A41604	myosin heavy chain
20	225.5	4.4	2094	2	S33124	tpr protein - huma
21	225	4.4	1017	2	PC4035	cell-cycle-depende
22	223	4.3	1286	2	T16507	hypothetical prote
23	222.5	4.3	1133	2	T22976	hypothetical prote
24	222	4.3	876	2	A23767	myosin heavy chain
25	222	4.3	2649	2	A40937	bullous pemphigoid
26	219	4.3	1940	2	A59287	myosin heavy chain
27	218.5	4.2	1388	2	S74245	serine/threonine-s
28	218.5	4.2	1955	2	T30934	myosin-like protei
29	218.5	4.2	3187	2	JC5837	364K Golgi complex
30	217.5	4.2	1976	2	A59252	myosin heavy chain
31	217	4.2	1938	2	JC5421	smooth muscle myos
32	217	4.2	1972	2	JC5420	smooth muscle myos
33	216.5	4.2	1388	2	S70633	serine/threonine-s
34	215.5	4.2	1034	2	T22166	hypothetical prote
35	215.5	4.2	1091	2	T34107	hypothetical prote
36	215	4.2	1837	2	T41023	probable nuclear p
37	215	4.2	1974	2	T30010	hypothetical prote
38	214	4.1	1407	1	S28589	trichohyalin - rab
39	213.5	4.1	1927	2	A59236	embryonic muscle m
40	213	4.1	2253	2	T30336	nuclear/mitotic ap
41	212.5	4.1	1940	1	A24922	myosin heavy chain
42	212	4.1	1957	2	A59294	skeletal myosin -
43	212	4.1	1957	2	T38077	hypothetical coile
44	212	4.1	2954	2	T14156	kinesin-related pr
45	211.5	4.1	1218	2	T14265	golgin-245 - mouse

ALIGNMENTS

```
RESULT          1
A46431
tight junction-associated protein ZO-1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A46431
R;Itoh, M.; Nagafuchi, A.; Yonemura, S.; Kitani-Yasuda, T.; Tsukita, S.; Tsukita, S.
J. Cell Biol. 121, 491-502, 1993
A;Title: The 220-kD protein colocalizing with cadherins in non-epithelial cells is id
py.
A;Reference number: A46431; MUID:93252986
A;Accession: A46431
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-1745 <ITO>
A;Cross-references: GB:D14340; NID:g303709; PIDN:BAA03274.1; PID:d1003784; PID:g30371
A;Experimental source: F9 cells
A;Note: sequence extracted from NCBI backbone (NCBIN:131200, NCBIP:131201)
C;Superfamily: guanylate kinase homology; GLGF domain homology
F;27-106/Domain: GLGF domain homology <GLG1>
F;428-498/Domain: GLGF domain homology <GLG3>
F;645-794/Domain: guanylate kinase homology <GKI>

Query Match            5.6%;       Score 289;  DB 2;  Length 1745;
Best Local Similarity  22.3%;      Pred. No. 2.5e-07;
Matches 191; Conservative 136; Mismatches 319; Indels 212; Gaps 43;
```

```
Qy   231 QELQRANMVSSCELELEQSLRTASDQESGDEELNRLKEENEKLRSLITFSLAEKDILEQS 290
         ::::||
Db    70 QENDRVAMVNGVSMDNVEHAFAVQQLRKSGKNA--KITIRRKKKVQIPVSHPDPEPVSDN 127

Qy   291 LDEARGSRQELVERIHSLRE-RAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNA 349
         ::::||        |  |:|   | :| |
Db   128 EDDSYD------EEVHDPRAGRGALANRRSEKSWARDRSA--SRERSLSPRSDRRSVAS 178

Qy   350 LQ---AQVCELQKERDQAYSARDSAQ--REISQ-SLVEKDSLRRQ---VFELTDQVCE 398
         :       |: ||| ||  :|||: || ||   ::|||:|| :|:| :|
Db   179 SQPAKPTKVTLVKSRKNEEYGLRPASHIFVKEISQDSLAARDGDIQEGDVLKINGTVTE 238

Qy   399 --LRTQLRQLQAEPPGVLKQEARTREPCPREKQRLVR-----MHAICPRDSDCSLVS 449
            ||:|  :| |::|||  |:|| || |    | || :|     : |||  | |:|||
Db   239 NMSLTDAKTLIERSKGKLKMVVQR-----DERATLLNVPDLSDSIHSANASERDDISEIQ 293

Qy   450 STESQLLSDLSATS----------SRELVDSFRSSSPAPPSQQSLYKRVAEDFGEE 495
         :     |     |::          :|||  :: || | | |: :||:|| |:|
Db   294 S----LASDHSGRSHDRPPRRSQSRSPDQSREPSDHSTQSPQQPSNGSLRSR-EEERMSK 348

Qy   496 PWSFSSCLE---------------IPEGDPGALPGAKAGDPHLDYELLDTA 531
         |||:     |               |:|  :| |    | | || | ||: :||
Db   349 PGAISTPVKHVDDHPPKAVEEVTVEKNEKQTPLPEPKP---VYAQVGQPDVDL------ 399
```

```
Qy   532  DLPQLESSLQPVSP--GRLDVS--ESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIG  587
               ||:|            ::  |||||||||||:|||:|
Db   400  -------PVSPSDGALPNSAHEDGIL--RPSMKLVK-----FRKGDSV--GLRLAG      439

Qy   588  GNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTL--------EEA  639
          ||  ||| || |         :  |                              | |
Db   440  GNDVGIFVAGVLEDSPAAKEGLEEGDQILRVN---NVDFTNIIREEAVLFLLDLPKGEEV  496

Qy   640  VGLLRRVDGFCCLSVKNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHC  699
           ||       |  :   ||||||    ||||     :|||||||||:||||||| | :
Db   497  TILAQK-----------KKDVYRRIVE------SDVGDSFYIRTHFEYEKESPYGLSFNK  539

Qy   700  NEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTV  758
           || || || | |   || | :|  | | :  |  ||||: |||:|||:|| :|
Db   540  GEVFRRVVDTLYNGKLGSWLAIRIGKNHKE-VERGIIPNKNRAEQ--LASVQ-------Y  588

Qy   759  TRKPSSGGPQKLVRIVSMDKAKASPLR-LSFDRGQLDPSRMEGSSTCFWAESCLTLVP-Y  816
          | | :||    |  ::     ||:|| |   | |: :       |  :  | : : |  |
Db   589  TLPKTAGG-----DRADFWFRGLRSSKRNLRKSREDLS----AQPVQTKFPAY        633

Qy   817  TLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEE---YE-AWSQRGD  872
               |  |  || :|  |  || |:                |  ::::   |:  | | :|
Db   634  ERVVLREAGFLRPVTIF-GPIADVAREK-------------LAREEPDIYQIAKSELRD   678

Qy   873  IIQEGEVSGGRCWTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMA  932
          |||| :      | : :   |  ||||||||| | :| ||||| :| |  :: ::|
Db   679  AGTDHRSSG---IIRLHTIKQIIDDQKHALLDVTPNAVDRLNYAQWYPIVFLNPDSKQG  735

Qy   933  KKL-------KKGLQRLGTSEEQLLEAARQEEGDLDRAPCLYSSLAPDGWSDLDGLLS   983
          ||        ||  :        :            :                 ||
Db   736  VKTMRMRLCPESRKSARKLYERSHKL----RKNNHHLFTTINLNSM-NDGW------YG  784

Qy   984  CVRQAIADEQKKVVWTEQ  1001
           :||||:::: :: :|
Db   785  ALKEAIQQQQNQLVWSE   802
```

RESULT 2
T13703
tamA protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13703
R;Takahisa, M.; Togashi, S.; Suzuki, T.; Kobayashi, M.; Murayama, A.; Kondo, K.; Miyake,
Genes Dev. 10, 1783-1795, 1996
A;Title: The Drosophila tamou gene, a component of the activating pathway of extramacroc
A;Reference number: 217700; MUID:96312452
A;Accession: T13703
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1367 <TAK>
A;Genetics:
A;Cross-references: EMBL:D83477; NID:g1498136; PIDN:BAA11923.1; PID:g1498137
A;Gene: tamou (tam)
A;Cross-references: FlyBase:FBgn0003177

```
Query Match             5.5%;  Score 282;  DB 2;  Pred. No. 4.3e-07;  Length 1367;
Best Local Similarity  22.7%;
Matches 155;  Conservative 111;  Mismatches 276;  Indels 142;  Gaps 30;

Qy   373  REISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLV  432
          :|||       |||:| :|  |:  |  :|    |
Db   194  KEISSK--AREQLNANGYSL-QEDIITRIHNTNCGDTMSLKEAKKIIDGC-KERLNLV   248

Qy   433  RMHAICPRDD-SDCSLVSSTESQLLSDLSAT----------SSRELVDSFRSSSPAPPSQ  481
          | | :  |      || ::: ||| |||               |||:|  |||  | :
Db   249  VLRDITNQTAVSQLNLNNSASHQASGNIYATHQPQVSGCSSSNNNLEDPYLPGG-ASYSS  307

Qy   482  QSLYKRVAEDFGEEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQ  541
          |:||            ||     ||| ||  |||||||||| |: :  : :|| |::| :
Db   308  QNLYVQPPTRTSNGPNINGNGLN--DEKSNLTPRGRSRGP------IMDGVSLQQLD---R  357
```

RESULT 3
A47747
tight junction protein ZO-1 - human
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C;Accession: A47747
R;Willott, E.; Balda, M.S.; Fanning, A.S.; Jameson, B.; Van Itallie, C.; Anderson, J.
Proc. Natl. Acad. Sci. U.S.A. 90, 7834-7838, 1993
A;Title: The tight junction protein ZO-1 is homologous to the Drosophila discs-large
A;Reference number: A47747; MUID:93361541
A;Accession: A47747
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1736 <WIL>
A;Cross-references: GB:L14837; NID:g292937; PIDN:AAA02891.1; PID:g292938
C;Superfamily: guanylate kinase homology; GLGF domain homology
C;Keywords: alternative splicing; membrane protein; phosphoprotein
F;15-94/Domain: GLGF domain homology <GLG1>
F;181-248/Domain: GLGF domain homology <GLG2>
F;416-486/Domain: GLGF domain homology <GLG3>
F;633-782/Domain: guanylate kinase homology <GKI>

```
Query Match             5.3%;  Score 270.5;  DB 2;  Length 1736;
Best Local Similarity  21.4%;  Pred. No. 2.4e-06;
Matches 174;  Conservative 129;  Mismatches 278;  Indels 233;  Gaps 38;

Qy   542  PVSPGR-----LDV------SESGVLM----------RRRPARRILSQVTMLAFQGDALL  580
           | ||            :|:      |||||||||||:|||:|
Db   358  PVTPTRGRSAAIDEPPRPPPRGSSSGAAQEDFYSSRRQLYEERQSAEPRFISFQKEGSV   417

Qy   581  EQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTL----  640
           :|||||||||||||:|| |||||||| :||: :                        |
Db   418  -GIRLTGGNEAGIFVTAVQPGPSPASLQGLMPGDKILKVN----------DMDMNGVTREEAV  468

Qy   641  GLLRRVDGFCCLSVKNTDGYKRLLQDLEAKVATS-GDSFYIRVNLAMEGRAKGELQVH  698
           | :        :   || :    :: :   |  |||| :| :  : :   | |
Db   469  LFL------LSLQDRIDLIVQYCKEEYDEVVTNQRGDSFHIKTHFCDNPSKGEMAFK    520

Qy   699  CNEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCT  757
           | ||:| || |||   || | |::|  |  |   |||||:  ||:|||:|| :|
Db   521  AGDVFRVIDTLHNGVGSWQVLKIGRGHQE--MQRGVIPNKSRAEEL--ATAQFN       571

Qy   758  VTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSR--------MEGSSTCFWAESCL  811
           | | :||                   | :       |:         |  :  | :
Db   572  -ATKK------EMNANESRGNFRRRRSTHRRSKSLSRENWDDVVFSDSIS            615

Qy   812  TLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRG  871
             |  | | |  |  || :|  |  || |:               | ::::  |:  | |
Db   616  KFPAYERVVLRHPGFVRPVVLF-GPVSDLARERLA--KDFPDKFSTPLQDDDKSA       667

Qy   872  DIIQEGEVSGGRCWTRHA-VESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEK  930
           :|||| :      | : :    |  ||||||||| | :| ||||: :| | |::: :|
Db   668  -----ATSGKCRIVRLSNIRDVMDRGKHALLDITPNAVDRLNYAQFYPVVVIFLKTDSK   720

Qy   931  -MAKKLLKKGLQRLG-TSEEQLLEAARQEEGDLDRAPCLYSSLAPDGW---------SDL  978
                   :|  :        :                                   ::|
Db   721  HVIKQLRHGLPKAAHKSSKKLLEQCQK----LERV-----------WSHIFSTQIALSDE  765

Qy   979  DGLLSCVRQAIADEQKKVVWTEQS  1002
            | : :||||:::: :: :|:
Db   766  ESWYRKLRDSIDLQQSGAVWMSES  789

Qy   247  QEQSLRRTASDQESGDEE-------------LNRLKEENEKLRSITFSLAEKDILEQSLDEARG  296
           :| |: :: :   ::               |||:|::|| |||::| |||||:||:|  |||
Db   151  RERSLSPRSDRRSVASSQPAKPTKVTLVKSRKNEEYGLR-LASHIFVKEISQDSLAARDG     209

Qy   297  SRQE--LVERIH-SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQ  353
           ::|    | :    | | :        :: :: :: :         :   :: |||| |
Db   210  NIQEGDVVLKINGTVTENMSLTD--------AKTLIERSKGKLKMVVQRDE----RAT    255

Qy   354  VCELQKERDQAYSARDSAQREIS--QSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPP  411
```

```
R;Harborth, J.; Weber, K.; Osborn, M.
EMBO J. 14, 2447-2460, 1995
A;Title: Epitope mapping and direct visualization of the parallel, in-register arrang
A;Reference number: S55331; MUID:95300777
A;Accession: S55331
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 247-279 <HAR>
R;Yang, C.H.; Lambie, E.J.; Snyder, M.
J. Cell Biol. 116, 1303-1317, 1992
A;Title: NuMA: an unusually long coiled-coil related protein in the mammalian nucleus
A;Reference number: S23647; MUID:9217231
A;Accession: S23647
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-771,'Q',773-814,'ER',817-872,'E',874-1267,'RLRLLQAETASNSARAAERSSALREEVQ
A;Cross-references: EMBL:Z11583; NID:g35118; PIDN:CAA77669.1; PID:g35119
A;Note: the authors translated the codon GAG for residue 781 as Gly, TTC for residue
C;Genetics:
A;Gene: GDB:NUMA1; NUMA
A;Cross-references: GDB:137229; OMIM:164009
A;Map position: 11q13-11q13
C;Keywords: mitosis; nucleus


Query Match             5.1%;  Score 262;  DB 2;  Length 2101;
Best Local Similarity   19.9%;         Pred. No. 8.7e-06;
Matches  208;  Conservative 163;  Mismatches 376;  Indels 296;  Gaps  39;

Qy      4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL  63
            ||  | :                  | |||      ::  :  |    | ||   :| |
Db    263 LLNEKQAASPLEPKELEELRDKN---------ESLTMRLHETLKQCQDLKTEKSQMDRK 312

Qy     64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS 122
                           :|                  ||| ::  |     : | ::|
Db    313 INQLSE------ENGDLSFKLREFASHLQQLQDALNELTEEHSKATQEWLEKQA 360

Qy    123 KLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182
           ||  ||    ||   ||   |  ||     :|  |  |             :
Db    361 QLEKELSAA--LQDKKCLEE-KNEILQGKLSQLEEHLS---------QLQDNPPQEKG 406

Qy    183 EVSAHFHEVLRLKDEMLSLS---------LHYSNALQEKELAASRCRSLQEELYL-- 228
           |:  | |          :|         :||   :| |  |        | |:|
Db    407 EVLGDVLQLETLKQEAATLAANNTQLQARVEMLETERGGQEAKLLAERGHFEEEKQQLSS 466

Qy    229 --------LKQELQRANMVSSCELELQEQSLRT------ASDQESGDEELNRLK 268
                    |:||||  :   |||||: ||:|        ||::|  | ||| ||
Db    467 LITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELTTLNATIQQQ-DQELAGLK 525

Qy    269 EE----------------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
           ::                      || | :   | | |  | ||: ||  :|| | :||
Db    526 QQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQK--EQQLKEV-AEKQEATRQDH 581

Qy    307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS 366
           :| || : |   ||:  |  : |:| |  |    |::||    : |  |  |::   |
Db    582 A-QQLATAAE-EREASLRERDAALKQLE---ALEKEKAAKLEI---LQQQLQVANE 629

Qy    367 ARDSAQREISQSLVEKDSLRRQVFEL--------------TDQVCELRTQLRQLQAEPPG 412
           ||||||  : :|  |:| ||  |  |              |||  |||: ||  ||
Db    630 ARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQ----- 684

Qy    413 VLKQEARTREPCPREKQRLVRMHAICPRDSDCSLVSSTESQLLSDSLSATSSRELVDSFR 472
            :|||||  || ||||||||| |||||||||||||  |||||| |||:|||||||:|:||
Db    685 --QKATEKERVAQEKDQL-------QEQLQALKESLKVTKGSLEEEKR 723

Qy    473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
           |||| |||||||||||||||||||||||||||||||||||| | : |||||||||||||
Db    724 RAADALEEQQRCISELKAE-----TRSLVEQHKRERKELEEERAGRKGLEARLLQLGE 776

Qy    533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
           |||||||||||||||||||||||||||||||||||||||||  ::  |||||||||| ||
Db    777 AHQAETE----------------------VLRRELAEAMAAQHT----AESECEQL----- 806
```

```
Db    256 LLNVPDLSDSIHSANASERDDISEIQSLASDHSGRS----------HDRPP 296
            :      :  ||  ||          |  ||  || ||     ||
Qy    412 GVLKQEARTREPCPREKQRLVRMHAICPRDSDCSLVSSTESQL----------LSDL 459

Db    297 ---RRSRSRSPDQRSEPSDHSRHS--PQQPSNGSLRSRDEERISKPGAVSTPVKHADDH 350
           ::|  ||  | |  |   ||   |:   |    :| :  | :|
Qy    460 SATSSREL-VDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKA 518

Qy    351 TPKTVEEVTVERNEKQTPSLFEPKPVYAQV---GNQMWIYLSV--------390
             |  |                 | |:    :| | | :|
Qy    519 GDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDA 578

Db    391 ---HLMVSYL------IQIMKMGFL--RPSMKLVK-----FRKGDS 420
               |::|       || :| || :|  :||
Qy    579 LLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPELFKAVLEDTTL- 636

Db    421 V--GLRLAGGNDVGIFVAGVLEDSPAAKEGLEGDQILRVN---NVDFTNIIREEAVLFL 475
            :| | ||  :: :|| |  |:  |  :||  | :|:|      | |:|::||| |:|
Qy    637 ------EEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGR 690

Db    476 LDLPKGEEVTILAQK----------KKDVIRRIVE------SDVGDSFYIRTHFEYEKE 518
            ||  | |  :|              ||||:|:||     | |  |||||| || | |
Qy    691 AKGELQVHCNEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALI 749

Db    519 SPYGLSFNKGEVFRAVDTLYNGKLGSWLAIRIGKNHKE--VERGIIPNKNRAEQ--LASV 574
           | |||||||:||| | | | |  |||||||||| | ||  |||||||||:|:|   ||:|
Qy    750 QDMTQQCTVTRKPSSSGGPQKLVRIVSMDKAKASPLR-LSFDRGQLDPSRMEGSSTCFWAE 808

Db    575 Q----------YTLPKTAGG-----------DRADFWRFRGLRSSKRNLRKSREDLS----AQ 612
           |          ||| : |||             ||||| || | | ||   || ||    ||
Qy    809 SCLTLVP-YTLVWPHRPARPRPVLLVPRAVGKILSEKICLLQGFKKCLAEYLSQEE---Y 864

Db    613 PVQTKFPAYERVVLREAGFLRPVTIF-GPIADVAREK----------LAREEPDIY 657
            | |:  |   | ||  | |||  |  |||| |             | ||| ||:
Qy    865 E-AWSQRGDIIQEGEVSGGRCWTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVI 923

Db    658 QIAKSEPRDAGTDQRSSG---YIRLHTIKQIIDQDKHALLDVTPNAVDRLNYAQWYPIVV 714
           |:| ||| ||| |||      ||||| |||| ||| |||||| | ||||| ||| ||||
Qy    924 HVSVNEKM-----------AKKLKKGLQRLGTSEEQLLEAARQEEGDLDRAPCLY 967

Db    715 FLNPDSKQGVKTMRMRLCPESRKSARKLYERSHKLAKNNHHLFTTT----------IN 762
                     |:|: |||||  | | |:|||    | ::     ||            |:
Qy    968 SSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQ 1001

Db    763 LNSMNDGW------YGALKEAVQQQQNQLVVVSE 790
            :|  | |      ||||||||: |::|::|||
Qy    763 YGALKEAVQQQQNQLVVVSE 790
```

RESULT   4

```
A42184
nuclear mitotic apparatus protein NuMA - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-2000
C;Accession: A42184; S33376; S55331; S23647; S24554
R;Compton, D.A.; Szilak, I.; Cleveland, D.W.
J. Cell Biol. 116, 1395-1408, 1992
A;Title: Primary structure of NuMA, an intranuclear protein that defines a novel pathway
A;Reference number: A42184; MUID:92176238
A;Accession: A42184
A;Molecule type: mRNA
A;Residues: 1-2101 <COM>
A;Cross-references: EMBL:Z11584; NID:g35120; PIDN:CAA77670.1; PID:g35121
A;Note: sequence extracted from NCBI backbone (NCBIN:85755, NCBIP:85760)
R;Tang, T.K.; Tang, C.C.; Chen, Y.L.; Wu, C.W.
J. Cell Sci. 104, 249-260, 1993
A;Title: Nuclear proteins of the bovine esophageal epithelium. II. The NuMA gene gives r
A;Reference number: S33376; MUID:93280231
A;Accession: S33376
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1705-2101 <TAN>
A;Cross-references: EMBL:214229; NID:g296118
A;Note: this translation is not annotated in GenBank entry HSNUMAT3G, release 113.0
```

```
Qy   593  IFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL  652
          :||  |||  |  | :      |    :|| :|  | ::  :|
Db   807  --------------VKEVAAWRDGYE----DSQQEEAQYGAMFQEQ------------------LM  836

Qy   653  SVKNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG  712
          : || :|    :|    |:    ::|| |:|  |  :| || |  |  ||:|
Db   837  TLKEECEKARQELQEAKEKVA--GIESHSELQISRQQNKLAEL--HAN-----------  880

Qy   713  CGGWHAHRVNSYTMKDTAAHGTIPNYSRAQQOLIALIQDMTQQCTVTRKPSSGGPQKLV-  771
          :  |:  :| :|  ::   |    |  |     ||  :| | :|| ||  ||
Db   881  ----LARALQQVQEKEVRAQKLADDLSTIQEKMAATSKEVARLETLVRK--AGEQQETAS  934

Qy   772  RIVSMDKAKASPLRLSFDRGQLDPSRME--GSSTCFWAESCLTLVPYTLVWPHRPARPR  828
          | |       :|  |::    | |:|     |:|
Db   935  RELVKEPARA-------GDRQPEWLEEQQGRQFC-----------------STQA  965

Qy   829  PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEYEAWSQRGDIIQEGEVSSGGRCWVTR  888
          | |||| ::|  | |:| :||  | |||| || |||| |||||  |:|  |   |
Db   966  ALQAMEREAEQMGNE---LERLRAALMESGQGOQQEERGQQQEREVARLTQERG------R  1015

Qy   889  HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI--VIHVSVNEKMAK-----KLK------  936
          | |  |:      |:    : | : | | |       |:|  ||  |||
Db  1016  AQADALALEKAAARAELEMRLIQNALNEQRVEFATIQEALAHALTEKEGKDQELAKLRGLEAA  1075

Qy   937  --KGLQRLGTSEEQLLEAARQEE  957
            :|||  | :  :||:|||
Db  1076  QIKELEELRQTVKQLKEQLAKKE  1098
```

RESULT   5

G02520

plectin - human

C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999
C;Accession: G02520
R;McLean, W.H.I.; Smith, F.J.D.
submitted to the EMBL Data Library, March 1996
A;Reference number: H01385
A;Accession: G02520
A;Molecule type: mRNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-4574 <MCL>
A;Cross-references: EMBL:U53204; NID:g1477645; PIDN:AAB05427.1; PID:g1477646
C;Genetics:
A;Gene: PLEC1
C;Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S
F;68-283/Domain: alpha-actinin actin-binding domain homology <ACT>

```
Query Match      4.8%;     Score 249;    DB 2;      Length 4574;
Best Local Similarity   22.4%;    Pred. No. 0.00012;
Matches 187;   Conservative 112;   Mismatches 319;   Indels 218;   Gaps 34;

Qy   135  LQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRL  194
          ||||   ||:|::    |    |  ||  |   |   |  |     |:
Db  2293  LQEEAEKMKQVAEEAARLSVAAQEARARLQLAEEDLAQQORALAEKMLKEKMQAVQEATRL  2352

Qy   195  KDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQEL--QRANMVSSCELELQEQ--  249
          |    |||  :    |:| |  | |  | |||| | |:   || ::| || |  :
Db  2353  ----QQKELAQEQARRLQEDKEQMAQQLAEETQGFQRTLEAERQRQLEM  2405

Qy   250  SLRTAS----------DQESGDEELNRLKEE-----NEKLRSLTFSLAEKDILEQSLDEAR  295
          ||  |:          |||  |:||| ::||     |::|:|:| | |   |:|: ||
Db  2406  SAEAERLKLRVAEMSRAQARAEEDAQRFRKQAEEIGEKLHRTELATQEKVTLVQTLEIQR  2465

Qy   296  GSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMAC----QLYREKVNALQ  351
          |||| :|:| |:|||   | || || |  | |     || |         | ||:| ||
Db  2466  QQSDHDAER---LREAIAELERERKEKL--QQEAKLLQLKSEEMQTVQQEQLLQETQALQQ  2520

Qy   352  AQVCE----LQKER----------DQAYSARDSAQREISQSLVEKDSL-----  385
          ||       |:| |          ||   |  | |  :|:| || |
Db  2521  SFLSEKDSLLQRERFIEQEKAKLEQLFQDEVAKAQQLREEQQRQQQMEQERQRLVASME  2580
```

```
Qy   386  --RRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDDS  443
            :||  |:         | :|||       :|   | |:|  |   :| : |
Db  2581  EARRRQHEAEEGVRRKQEELQQLE----QORQQEELLAEENQRLREQLQLLEEQH  2632

Qy   444  DCSLVSS---TESQLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFS  500
          | :| |    | || | | :|  | :|  ||      | :|| || ||   |  |||||
Db  2633  RAALAHSEEVTASQVAATKTLPNGRDALDGPAAEAEPEHSFDGLRRKVSAQRLQEAGILS  2692

Qy   501  SCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRR  560
          |  |   | |       |            :  || :   | :::| | ::  |||
Db  2693  A------------------------------EELQRLAQGHTTVDE---LARRE  2713

Qy   561  PARRILSQVTMLAFGQDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDY  620
          | |:| |:||   |:|| |:||        ::   | :| | ::|   ||  || | :|
Db  2714  DVRHYL-------QGR-----SSIAGLLLKATNEKLSVYAALQRQLLSPGTALILLEA  2759

Qy   621  EAS------EPLFKAVLEDTTLEEAV--GLLRRVDGFCCLSVKVNTDGYK---------  662
          :|       :|   ||: | |  | |  |  ::   :||  |  || ||
Db  2760  QAASGFLLDPVRNRRL--TVNEAVKEGVVGPELHHKLLSAERAVTGYKDPYTGQQISLF  2816

Qy   663  RLLQD--------LEAKVATSG----DSFYIRVNLAMEGRAKGELQVHCNEVLHVT  706
          | |          |  | |||      | |||||   |||   | | |  || ||
Db  2817  QAMQKGLIVREHGIRLLEAQIATGGVIDPVHSHRVPVDVAYR---RGYFDEEMNRVLADP  2873

Qy   707  DTMFQCGCGWHAHRVNSYTMKDTAAHGTIPNYSRAQQOLIALIQD-MTQQC--TVTRKPS  763
          ::    |     :|::        | :      | :|   | |    |: |    |:| |
Db  2874  SDDTKGFFDPNTHENLTYL-------------QLLERCVEDPETGLCLLPLTDKAA  2916

Qy   764  SGGPQKLVRIVSMDKAKASPLRLSFDRGQLDP--SRMEGSSTCFWAESCLTLVPYTLVWP  821
          :|||| :| |:|   |:|:|  |:    |:|    | | |  | ||   :|| ||:
Db  2917  EGG-------ELVYTD---SEARDVFEKATVSAPFGKFQGKTVTIW--EIINSEYFT----  2961

Qy   822  HRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEYEAWSQRGDIIQEG  877
          :| || ||:|    | |  |||::  |  :|  | |  : |  :     : | |
Db  2962  ---AEQRRDLLRQFRTGRITVEKII-------KIIITVVEEQE------QKGRLCFEG  3003
```

RESULT   6

A59404

plectin [imported] - human

C;Species: Homo sapien (man)
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C;Accession: C59404; A59404
R;Liu, C.G.; Maercker, C.; Castanon, M.J.; Hauptmann, R.; Wiche, G.
Proc. Natl. Acad. Sci. U.S.A. 93, 4278-4283, 1996
A;Title: Human plectin: organization of the gene, sequence analysis, and chromosome 1
A;Reference number: C59404; MUID:96210632
A;Accession: C59404
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-4684 <STO>
A;Cross-references: GB:CAA91196; NID:g1296662; PIDN:CAA91196.1
C;Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protei

```
Query Match      4.8%;     Score 249;    DB 2;      Length 4684;
Best Local Similarity   22.4%;    Pred. No. 0.00012;
Matches 187;   Conservative 112;   Mismatches 319;   Indels 218;   Gaps 34;

Qy   135  LQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRL  194
          ||||   ||:|::    |    |  ||  |   |   |  |     |:
Db  2403  LQEEAEKMKQVAEEAARLSVAAQEARLRQLAEEDLAQQORALAEKMLKEKMQAVQEATRL  2462

Qy   195  KDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQEL--QRANMVSSCELELQEQ--  249
          |    |||  :    |:| |  | |  | |||| | |:   || ::| || |  :
Db  2463  KAEAELLQ------QQKELAQEQARRLQEDKEQMAQQLAEETQGFQRTLEAERQRQLEM  2515

Qy   250  SLRTAS----------DQESGDEELNRLKEE-----NEKLRSLTFSLAEKDILEQSLDEAR  295
          ||  |:          |||  |:||| ::||     |::|:|:| | |   |:|: ||
Db  2516  SAEAERLKLRVAEMSRAQAREEDAQRFRKQAEEIGEKLHRTELATQEKVTLVQTLEIQR  2575

Qy   296  GSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMAC----QLYREKVNALQ  351
          |||| :|:| |:|||   | || || |  | |     || |         | ||:| ||
Db  2576                                              ---QLYREKVNALQ
```

```
Qy  229  LKQE------...LQRANMVSSCELELEQSLRTASDQESGDEELNRLKEENEKLRSLTFSL  281
         :||                          :|  |: |:  || : | : :| :| :
Db  559  LRQEQTEVTAALARAEQ-SIAELSSSSENTLKT------EVADLRAAAVKLSALNEAL    608

Qy  282  A-EKDILEQSL---DEARGSRQELVERI-HSLRERAVAAERQREQYWEKEQTLLQF      333
         |  : | :||     ||:     ||:   :|: ||  |  |::|:  :|::| ::||
Db  609  ALDKVGLNQQLLQLEEENQSVCSRMEAAEQARNALQVDLAEAEKRREALWEKNTHLEAQL  668

Qy  334  QKSKMACQLYREKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELT  393
         ||:       || :||  :|::: |||  |  ::|: |:   :|::
Db  669  QKAE------...EAGAELQADLRDIQEEKEE------IQKLSESRHQQEA          705

Qy  394  DQVCELRTQLRQLQAE--PPGVLKQEARTREPCCPREKQRL-VRMHAICPRDDSDCSLVS  449
         |  || |:||:| |:|       |||  |   |  |    |
Db  706  -----ATTQLEQLHQEAKRQEEVLARAVQEKEALVREKAALEVRLQAV-ERDRQDL---  755

Qy  450  STESQLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGD  509
         :   ||:  :: |||:||||                 |
Db  756  AAQLQGLS------...SAKELLES------SLF--------EAQQQNSVIDEPQG-      788

Qy  510  PGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGR----...LDVSESGVL         556
                                                  |: |:
Db  789  ------...QLEVQIQTVQAKEVIQGEVRCLKLELDTERSQAE                 823

Qy  557  MRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIV  616
         |||                  |:| :|
Db  824  QERDAAARQLAQAEQ--EGKTALEQQKAAHEKEVNQLREKWEKERSWHQQELAKALE--   878

Qy  617  MVDYEASEPLFKAVLEDTTLEEAVGLLRRVD------...GFCCLSVKVNTDGYKRL--LQDL  668
                                 ||  |:|          |  || :| | |:|
Db  879  SLEREKME-LEMRLKEQQTEMEAIQAQREEERTQAESALCQMQLETEKERVSLLETLLQT  937

Qy  669  EAKVATSGDSFY------...IRVNLAMEGRAKGELQVHCNEVLHVTDTMFQGCGCWHAHRVNS  723
                                |||:|||||||||||:|:||||||||:|||||||||||||||
Db  938  QKELADASQQLERLRQDMKVQKLKEQETTGLQTQLQE------...AQR---              978

Qy  724  YTMKDTAAHGTIPNYSRAQQQLIALIQD-MTQQCTV---...TRKPSSGGPQKLVRIVSMDK   778
         ::|  |                                  |||
Db  979  -ELKE-AARQHRDLAALQEESSLLQDKMDLQKQVEDLKSQLVAQDDSQRLVEQEVQEK    1036

Qy  779  AKASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVG  838
                              :|
Db  1037 LRETQ------...EYNRIQKELEREKASLT------...LSLM---                 1075

Qy  839  KILSEKLCLL------...QGFKKCLA---EYLSQE---...EYEAWSQRGDIIQEGEVSGGRC   884
         ||                 |                     |:  |:| |||  |||:   |
Db  1076 SIRQQELSALRQDMQEAQGEQKELSAQMELLRQEVKEKEADFLAQEAQLLEELEAS----   1131

Qy  885  WVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGT  944
                                                          |:
Db  1132 ------...HITEQQLRASLWAQEAKAAQLH------...LRLRS                    1158

Qy  945  SEEQL-LEAARQEEGDLDRAPC---...LYSSL------...APDGWS---               976
         :|||   ||||:||| |  |         | |:            | |:
Db  1159 TESQLEALAAEQQPGNQAQAQAQLASLYSALQQALGSVCESRPELSGGGDSAPSVWGLEP  1218

Qy  977  DLDGLLSCVR------...QAIADEQKKV---VWTEQSPR                      1004
         | :| |  :|        | |:|:|         |:|  |
Db  1219 DQNGARSLFKRGPLLTALSAEAVASALKLHQDLWKTQQTR                     1259
```

RESULT 8

I54378

gene X104 protein - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I54378
R;Duclos, F.; Rodius, F.; Wrogemann, K.; Mandel, J.
Hum. Mol. Genet. 3, 909-914, 1994
A;Title: The Friedreich ataxia region: characterization of two novel genes and reducti
A;Reference number: I54378; MUID:95038744
A;Accession: I54378
A;Status: preliminary; translated from GB/EMBL/DDBJ

```
Db  2576 QQSDHDAER--...LREAIAELERERKEKL-QQEAKLLQLKSEEMQTVQQEQLLQETQALQQ  2630
                    :| | :|||::||:| || ||  |: |: |::|   |:|:|:|
Qy  352  AQVCE------...LQKER------...DQAYSARDSAQREISQSLVEKDSL-----      385

Db  2631 SFLSEKDSLLQRERFIEQEKAKLEQLFQDEVAKAQQLREEQQRQQQMEQEQRLVASME    2690
                          ::| :| | |||: || :|   | :|::|  :  |::|:|:|
Qy  386  --RRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCCPREKQRLVRMHAICPRDDS  443

Db  2691 EARRRQHEAAEEGYRRKQEELQQLE----...QQRRQQEELLAEENQRLREQLQLLEEQH    2742
         :||||:|:|       :::||    |      | :|:|:|  ||:|:||:|:| |::|
Qy  444  DCSLVSS--...TESQLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFS    500

Db  2743 RAALAHSEEVTASQVAATKTLPNGRDALDGPAAEAEPEHSFDGLRRKVSAQRLQEAGILS  2802
                        ::|: || : |:        |  |   | :   |||: |||:|:  ||:
Qy  501  SCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRR  560

Db  2803 A------...EELQRLAQGHTTVDE--LARRE                             2823
                   :||:|| |   |:|:   |||
Qy  561  PARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDY  620

Db  2824 DVRHYL------...QGR------...SSIAGLLLKATNEKLSVYALQRQLLSPGTALILLEA  2869
                    |             ||:| :|| :||: :  |::|| |||| | ::|| :|::|||
Qy  621  EAS---...EPLFKAVLEDTTLEEAV---GLLRRVDGFCCLSVKVNTDGYK-----      662

Db  2870 QAASGFLLDPVRNRRL---...TVNEAVKEGVVGPELHHKLLSAERAVTGYKDPYTGQQISLF  2926
                ||| :|              ||| |   ||    :|:|| | ::|||| |:||:||| ||
Qy  663  RLLQD------...LEAKVATSG---DSFYIRVNLAMEGRAKGELQVHCNEVLHVT         706

Db  2927 QAMQKGLIVREHGIRLLEAQIATGGVIDPVHSHRVPVDVAYR--RGYFDEEMNRVLADP   2983
         | | |      ||:||||||||||| |||||||||| || ||  ||||||||||||||||
Qy  707  DTMFQGCGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQD-MTQQC--TVTRKPS  763

Db  2984 SDDTKGFFDPNTHENLTYL------...QLLERCVEDPETGLCLLPLTDKAA           3026
             | |||||||||||: |        ||| | ||||||| ||||||||||
Qy  764  SGGPQKLVRIVSMDKAKASPLRLSFDRGQLDP--SRMEGSSTCFWAESCLTLVPYTLVWP  821

Db  3027 KGG------...ELVTD---SEARDVFEKATVSAPFGKFQGKTVTIW--EIINSEYFT---   3071
                   |||    ||| | || |||||||||| || |||||||||||  |||||||||
Qy  822  HRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEG      877

Db  3072 ---AEQRRDLLRQFRTGRITVEKII------...KIIITVVEEQE-----QKGRLCFEG    3113
```

RESULT 7

T08621

centrosome associated protein CEP250 - human
C;Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: T08621
R;Mack, G.J.; Rees, J.; Sandblom, O.; Balczon, R.; Fritzler, M.J.; Rattner, J.B.
Arthritis Rheum. 41, 551-556, 1998
A;Title: Autoantibodies to a group of centrosomal proteins in human autoimmune sera read
A;Reference number: Z16462; MUID:98165428
A;Accession: T08621
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2442 <MAC>
A;Cross-references: EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1; PID:g2832237
A;Experimental source: cell line HeLa

```
Query Match          4.8%;  Score 248.5;  DB 2;  Length 2442;
Best Local Similarity 22.1%;  Pred. No. 5.5e-05;
Matches 221;  Conservative 140;  Mismatches 341;  Indels 299;  Gaps 47;

Qy  123  KLT---ECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLH------QL  173
         |||      |          ||:   |:|:| ||||:|||| |:|  |:
Db  439  KLTGERDTLAGQTVDLQOGEVDSLSKERELLQQAREELRQQLEVLEQEAWRLRRVNVELQL  498

Qy  174  EADHSR-MKREVSAHFH---EVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYL   228
         |   |  ||||               ||||::||:|| ||||||   ||  |||:|
Db  499  QGDSAQGQKEEQQEELHLAVRERERLQEMLMGLEAKQSESLSELITLREALESIHLEGEL  558
```

A;Molecule type: mRNA
A;Residues: 1-1116 <RES>
C;Genetics:
A;Cross-references: GB;L27476; NID:g498012; PID:g498013
A;Gene: X104
C;Superfamily: guanylate kinase homology; GLGF domain homology
F;37-116/Domain: GLGF domain homology <GLG1>
F;731-879/Domain: guanylate kinase homology <GKI>

Query Match   4.8%; Score 246.5; DB 2; Length 1116;
Best Local Similarity 21.2%; Pred. No. 2.5e-05;
Matches 212; Conservative 137; Mismatches 393; Indels 257; Gaps 47;

Qy   55 EEVLHSPRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNP--DVYTLVTGLQPDVDF 112
           ::|||| ||| |: ::   :| :| |:| | |
Db   94 EDVLHS-FAVQQLRKSGKVAAIVVKRPRKVQVAALQA----SPPLDQDDRAFEVMDEFDG 148

Qy   113 SNF-SGLMETSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLH 171
           | || || : | : | | :::|: |
Db   149 RSFRSGYSERSRLNS-HGGRSRSWED--SPERGRPH--ERARSRERDLSRDRSRGRSLE   202

Qy   172 Q-LEADHSRMKREVSAHFHEVLRLKDEMLSLSL-HYSNALQEKELAASRCRSLQEELYLL 229
           : : :|:| | : |: ::|| ::: |
Db   203 RGLDQQDHARTRD----------RSRGRSLERGLDHDFGPSRDRDRSRGRSID------ 246

Qy   230 KQELQRANMVSSCELELQEQSLRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQ 289
           |: :||
Db   230 -QDYERAY------------------------HRAYDPDYER---------------AY 265

Qy   290 SLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNA 349
           : ||| :|: | | :|:|
Db   266 SPEYRRGAR-------HDARSRG-PRSRSREHPSRSPSPEPRGRPGPIGVLLMKSRAN- 316

Qy   350 LQAAVCELQKERDQAYSARDSAQ---REISQS-LVEKDSLRRQ---VFELTDQVCE--LR 400
              | | :| | || :: || |
Db   317 ------------EEYGLRLGSQIFVKEMTRTGLATKDGNLHEGDIILKINGTVTENMSL 363

Qy   401 TQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDD--SDCSLVSSTES----- 453
            || || || | :: | ||:|::|: |: | |: |
Db   364 TDARKLIEKSRGKL-QLVVLRD----SQQTLINIPSLNDSDSELEDISEIESTRSFSPEE 418

Qy   454 --QLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPG 511
               | |: |:| || ::| | : ||| |::||:|: ||
Db   419 RRHQYSDYDYHSSSSEKLKERPSSREDTPSRLS--------RMGATPTPFKS-------TG 463

Qy   512 ALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRRPARRIL--SQV 569
             | :   ::   :   :|| :   : :| :::|||:|||||: : |:|
Db   464 DIAGTVVPETNKEPRYQEPPAP------QPKAAPR--------TFLRPSPEDEAIYGPNT 510

Qy   570 TMLAF-QGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFK 628
              | ||:|||||| | |:| |||||| ||| ||| || | :||:||||| |
Db   511 KMVRFKKGDSV--GLRLAGGNDVGIFVAGIQEGTSAEGQEGLQEGDQIILKVNTQDFRGL- 566

Qy   629 AVLEDTTLEEAVGLLRRVDG-FCCLSVKVNTDGYKRLLQDLEARKVATSGDSFYIRVNLAM 687
           || :::: |:| | ||: ||||||||||:| | :   :||| |||| :
Db   567 -VREDAVLY----LLEIPKGEMVTILAQSRADVYRDIL-----ACGRGDSFFIRSHFEC   615

Qy   688 EGRAKGELQVHCNEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLI 746
            || |||:| | |: | || || | || : ::| | : |: |: ||:|
Db   616 EKETPQSLAFTGRGEVFRVVDTLYDGKLGNWLAVRIGNELEK-----GLIPNKSRAEQ--M 668

Qy   747 ALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSSTCFW 806
           | |: || |||:| | |:: |:|||| | |:| ||: :|:||
Db   669 ASVQN------AQRDNAGDRADFWRM---RGQRSGVKKNLRKSREDLTAVVSVSTKFP 717

Qy   807 AESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEA 866
             ||| : | ::| | | : |: :||: ||||:|: ||:|:| ||
Db   718 A-------YERVLLREAGFKRPVVLF-GPIADIAMEKLA--------NELPD 753

Qy   867 WSQRGDIIQEGEVSGGRCWVTR-HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHV 925
           |||| | |: || | | | |:|| |:|| |:: | :| : | |
Db   754 WFQTAKTEPKDGAGSEKSTGVVRLNTVRQVIEQDKHALLDVTPKAVDLLNYTQWFSIVISF 813

Qy   963 APCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQ 1001
              :   :   :   : |
Db   861 ------NSANDSW------FGSLKDTIQHQQGEAVWVSE   887

RESULT 9
T47237
myosin II heavy chain [imported] - Naegleria fowleri (fragment)
C;Species: Naegleria fowleri
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47237
R;Shaw, D.R.; Sullivan, P.K.; Marciano-Cabral, F.; Ennis, H.L.
submitted to the EMBL Data Library, December 1995
A;Description: Codon usage in Naegleria fowleri.
A;Reference number: Z24413
A;Accession: T47237
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-746 <SHA>
A;Cross-references: EMBL:U43192; PIDN:AAB01786.1
A;Experimental source: strain LEE mp: cell type amoeba

Query Match   4.8%; Score 245; DB 2; Length 746;
Best Local Similarity 19.2%; Pred. No. 1.7e-05;
Matches 160; Conservative 151; Mismatches 293; Indels 228; Gaps 31;

Qy   47 KVLCQLDEEEVLHSPRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGL 106
           | : | :|::|||||||||||||||||||||||||||||||||||||| ||| |: |
Db   19 KKLSESSKDELTLQLNKTND--EKNELVNKLKKAEKD-----LKNLKKSKDD------L   64

Qy   107 QPDVDFSN------FSGLMETSKITECLAGAIGSLQEELNQEKGQK----EVLL---- 150
            |:: :| |||||| || || || ||:|
Db   65 QAEKDDSDNRIRKLEQDLREKEQLSENLAKRIADLENEARTKEAQKKSTEMELSSVKDDL 124

Qy   151 ----RRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRLKDEMLSLSLHYS 206
               ||| ||:: | || :| |:|:| |
Db   125 NRTKQRAEQLQSDLEAQRERANELENLLSDTEGGKNQLDSQFKQL-----------Q 170

Qy   207 NALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEQSLRTASDQESGDEEL-N 265
           ||: |:| | | ||:| |:   :|:: |: ||:| ::
Db   171 NELQNERTNLQKMKSENERL------------QRELEEMK-RSLSDKQNESTSLDS 213

Qy   266 RLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIHSL---RERAVAAERQREQY 322
           ||| |:|||||| ||||| |: || || || :|:::|: ||:| | |::|
Db   214 KVKSLEDKIRELTALLETERSSKTDLDKKRSKMDKEVKRLAQQLQETEQALKGETQKK-- 271

Qy   323 WEEKEQTLLQFQKSKMACQLYREKVN----------------------ALQAQVCEL 357
           |::|::|| ||| :| | ||:|
Db   272 -NDADNRVKQLESELQGVKSERDRLNKDLNNTSGDMNGLKRQLDESNNLVAKLKAEIQKL 330

Qy   358 QKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQ----LQAEP 410
           |:|:| |:| :|| | :: :: || : :| | | | :|
Db   331 QKDLSDHHGDREETEEQL------DALRKQLQELTSRLSDANQKTQQEAASRQNLESEN 383

Qy   411 PGVLKQEARTREPCPREKQRLVR-MHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVD 469
              ||:|| | ::| :: ||:   :   :::|
Db   384 NRLKSEVSRLREDLQNENRRLKQEMERVQSESENE------KSELLTQLQ-------- 427

Qy   470 SFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLD 529
           | | : || | : |:   : | ||| :: |:
Db   428 ----------KLQEAYSEVKDELKD----LSKNASRGGGVVGGVDSAEVEKLRREYEM-- 471

Qy   530 TADLPQLESSLQPVSPGRLDVS--------------------ESGVLMRRRPARRILSQVTMLAFQG 576
            | | ||::|| : ||| :   :| |: ||| |:   :| |
Db   472 --QLAQLKARVEEVTQQRVDVENKKRSVEMDLTEMKTRLQTEERLRKKVEQQKSVEMEC 529

Qy   577 DALLEQISVIGGNLTGIFIFHRVTPGSAADQMALRPGTQIVMVDYEA-SEPLFKAVLED-- 633

```
Db   530  ---------------------------LAEEAEDLRDELNRTKLEHQALIQQLRQDLLQERH  569
                                      :::|.|         | |    ||:::|:
Db  1053  MQQAASLLMLQGGCQATAGEAEGDGAMSLLQQGEQLLEENGDVLISLQRAHEHAVKENA  1112

Qy   634  --TTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRA  691
             .|||:: |||| |||         : ||
Db   419  RTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESOLLSDLSATSSRELVDSFRSSSPAP  478

Db   570  SRASAEESATRQKR---------------EIEELQQDLEQERA--------KLDEAA  603
             : |           |||||:|||              ||||
Db  1113  KMATEIYQLQQRLKKL-------------EPGSVISSCLEEGTSEISG-SSREQVEPIMKQGPAT  1163

Qy   692  KGELQVHCNEVLHVTDTMFQGGCGCWHAHRVNSYTMKD-TAAHGTIPNYSRAQQLIALIQ  750
             |       |.||                    |||:::|
Db   479  PSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPG--------------AKAGDPHLD-YEL  527

Db   604  RRLKQQYENEILDLNNQIAQ-------AKKERSAASRDMKKADRDLREYQRRFQEEARAKQ  657
                        |              |                  |:|
Db  1164  ------KHFLSDLGDH------------------EARDLASTGTSSVQRQKCKTEASASLDCFSE  1205

Qy   751  DMTQQCTVTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSS  802
             |.|:::|                  | |:|                :||
Qy   528  LDTADLPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDA-----LL  580

Db   658  DLEQRLIKVEREN------KLLQ--SQSQSDAS----KYQKAEQEKQRLEAEN  698
             |            |     :::     :::         |:| |  | ||
Db  1206  LENSEDTRTESDL---------------KSQISQLREQLTVLRADCDRASERKQDLL  1247

Qy   581  EQISVIGGNLTGIFIHRVTPGS--------------------AADQMALRPGTQIVMVDYEASEPLFKAV  630
             :||                                        :::          |: :|
Db  1248  FDISVLKKKLK--MLERLPEASSRYKVLYEDAARENSCLQEELRLVETRYEESLDSNKEL  1305

Qy   631  ------LED------TTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKV-  672
                           :|:
Db  1306  TAEVRLQDEMKKMEEVMETFLSLEKSYD-----EVKVENEELRALVLRLQGKMEKVLGR  1360

Db   673  -ATSGDSFYIRVNLAMEGRAKGE-LQVHCNEVL--------------HVTDTMFQG  712
                ||                :| :| |
Db  1361  AALQGDSY---------ASGKPPSENLEVASDEKMLELRQTPKECTPKVVSMHHIIEBCTQE  1413

Qy   713  CGC------------------WHAHRVNSYTMKDTAAHGTIPNYSRAQQOLLIALIQD--M  752
             |                     ::  |
Db  1414  TQCCEQGSTKLLARIKAHEIAWFHRAIKTHPEKPSSAQNRVIPEGSAA----LLGL-QDKHL  1469

Qy   753  TQQCTVTRKPSSGGPQKL------------VRIVSMDKAKASPLRLSFDRGQLD--PSRMEGS  801
             |                             :||
Db  1470  QQEATIAELELE--KQKLQELTRNLRERVTALVRQKDAP-----SGQKEEELKAMMQPA  1522

Qy   802  STCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQ---------GF  851
             |                                                         |
Db  1523  VTCGEMQRKVELLRYE---------------------------SEKLQEENSILRNETTLNEEDSI  1562

Qy   852  KKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVC  911
             |:    |                                    |:
Db  1563  SNLKLEELNGSQEELWQKIETIEQE----------------KASIQTMVEKLKKQVSDLKIKN--  1609

Qy   912  TLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGTSEEQLLE--AARQEEG----------  958
             |                                     :||
Db  1610  --QQLD---------SENIELSQKNSQNKEELKTLNQRLAEMLCQREEPGACTSEKEWEQEN  1659

Qy   959  -----DLDRAPCLYSSLAPDGWSDLDGLLSCVR--QAIADEQKKVVWTEQSPR  1004
             |
Db  1660  ASLKEELDHYKVQTSTLV----SSLEAELSRIKLQTHVMEQENLLLKDELER  1707
```

```
RESULT  10
T30171
ninein - mouse
N:Alternate names:
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999
C:Accession: T30171
R:Bouckson-Castaing, V.; Moudjou, M.; Ferguson, D.J.P.; Mucklow, M.; Belkaid, Y.; Milon,
 J. Cell Sci. 109, 179-190, 1996
A:Title: Molecular characterisation of ninein, a new coiled-coil protein of the centroso
A:Reference number: 220751; MUID:96431720
A:Accession: T30171
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2168 <BOU>
A:Cross-references: EMBL:U40342; NID:g1113864; PIDN:AAA83234.1
                    EMBL:U40342; NID:g1113865; PIDN:AAA83234.1
A:Experimental source: strain C57B1/6
A:Note: localised specifically in the pericentriolar matrix of the centrosome
```

```
Query Match      4.7%;  Score 243.5;  DB 2;  Length 2168;
Best Local Similarity  19.6%;  Pred.No. 8.6e-05;
Matches  246;  Conservative 183;  Mismatches 380;  Indels 443;  Gaps 61;
```

```
Qy    13  ALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRLTNSAMRAGH  72
             ::: ||                        |       |:
Db    72  ...

Db   639  SIEAELVIEQMEQHH--RDLCHLRL--------------ELEDKVRHYEKQLDDTRVASEQ  684

Qy    73  LLDLLKTRGKNGAIAFLESLKFHNPDVYTL--VTGLQPDV-DFSNFSGLM---------119
             |:||||:                       :| |:
Db    73  ---------------------VHTLEKRVSELRSEIADLEGQAAVLREAHHKASC  730

Qy   685  EQAAMKQKYEQG--------------------VHTLEKRVSELRSEIADLEGQAAVLREAHHKASC  730

Qy   120  ----ETSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQL-----QEHLGLA-----ET  165
             ||||                              ||||
Db   165  RHEEEKRQLQMAFDEEKRAQLQEELRQE--HERELQARLQQAAETFRQEREGLAQAAWTEE  788

Qy   166  RAEGLHQ------------LEADHSRMKREVSAHFHEVLRLKDEMLSLSLHYSNALQE-KELAA  216
             |
Db   216  ...

Db   789  KVRGLEQSYQEQLLSLEEKHALEKEE-------LREE--LSEHHRRELQEGREEME  835

Qy   217  SRCRSLQEELYLLKQELQRANMVSSCE--LELQEQSLRTASDQESGDEELNRLKEENEKL  274
             |||
Db   274  ...

Db   836  TECNRRVSQI-----------EAQCQADCEKVTEHCEQTLQSLEVRHR--QELRDLLDQHLEE  885

Qy   275  RS-LTFSLAEKDILEQSLDEARGSRQELVERIHSLRERAVAAERQREQ-------YWE----324
             || |||                              ||||
Db   324  ...

Db   886  RSQWEF----EKDELAQECTDAQEQLKEALQ-----RERATAAMKQEQEILERTYKDRLN  937

Qy   325  ----EKEQTLLQFQKSKMACQLYREKVNALQAQVCELQEKRDQAYSARDSAQ-------372
             ||||
Db   372  ...

Db   938  ILSTEREQLLQDLKDLQNASE--SQHGLLSGQILEL--KRSQERELRDQGQALCQTGVS  992

Qy   373  -------------------------REISQSLVEKDSLRRQVFELTDQ-----VCELRTQLRQ  405
             ||
Db   405  ...

Db   993  EQLASPELERLRVEHEQERREMTCKLAALESAHRASLERADQEKAEMSTEICRLQNTVKD  1052

Qy   406  LQ---------------AEPPGVL--------------------------KQEA  418
             :|                   ||  ||                       | |
```

```
RESULT  11
I52300
giantin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C:Accession: I52300
R:Sohda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
 Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized
A:Reference number: I52300; MUID:95100974
A:Accession: I52300
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: GB:D25542; NID:g662389; PIDN:BAA05025.1; PID:g808869
C:superfamily: giantin
```

```
Query Match      4.7%;  Score 242.5;  DB 2;  Length 3225;
Best Local Similarity  19.3%;  Pred.No. 0.00017;
Matches  220;  Conservative 194;  Mismatches 396;  Indels 331;  Gaps 49;
```

```
Qy   954 RQEEGDLDRAPCLYSSL----------APDGWSDLDGLLSCVRQAIADEQKKVVWTE 1000
         |||  ||       ::   :  :  :::  :  ||:|::  :  : :|:|:
Db  1606 RQEKQE-----------LYGKLRSTEANKKETEKQLQEAEQEMEEMKEKMRKFAKSQKQKILELE 1659

Qy  1001 Q 1001
           |
Db  1660 E 1660
```

```
RESULT  12
A56539
giantin - human
N;Alternate names: macrogolgin
C;Species: Homo sapiens (man)
C;Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
R;Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
A;Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protei
A;Reference number: A56539; MUID:94187728
A;Accession: A56539
A;Molecule type: mRNA
A;Residues: 1-3259 <SEE>
A;Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715
C;Genetics:
A;Gene: GDB:GOLGB1; GCP; GCP371
A;Cross-references: GDB:454958
A;Map position: 3q13.31-3q13.31
A;Superfamily: giantin
C;Keywords: coiled coil; Golgi apparatus; transmembrane protein
F;3238-3254/Domain: transmembrane #status predicted <TMN>
```

```
Query Match             4.7%;    Score 242.5;   DB 1;    Length 3259;
Best Local Similarity  19.3%;    Pred. No. 0.00017;
Matches  220;    Conservative  194;    Mismatches  396;    Indels  331;    Gaps  49;
```

```
Qy    9 SALTALDEE---TLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLD-EEVLHSPRLT 64
Db  737 SAFTALSEERDQLLSQVKELSMVTELRAQV-KQLEMNLAEAERQRRLDYESQTAHDNLLT 795

Qy   65 NSAMRAGHLLDLLKTRGKNGAIAFLES------LKFHNPDVYTLVTGLQPDVDFSNFSGL 118
Db  796 EQI----HSLS-IEAKSKDVKIEVLQNELDDVQLQFSEQS--TLIRSLQSQLQ------- 841

Qy  119 METSKLTECLAGA-----IGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETR--AEGLH 171
Db  842 ---NKESEVLEGAERVRHISSKVEELSQALSQKELEITKMDQL----LLEKKRDVETLQ 893

Qy  172 QLEADHSRMKREVSAHFHE-VLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLL- 229
Db  894 QTIEEKDQQVTEISFSMTEKMVQLNEEKFSLGV----------------EIKTLKEQLNLLS 939

Qy  230 ------KQELQRANMVSS-CELELEQSLRTASDQESGDEELNRLKEENEKLRSLTFSLA 282
Db  940 RAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKEELQHEFDLLKKENEQRK------- 992

Qy  283 EKDILEQSLDEARGSRQELVERIHSLRERAV----------AAERQREQYWEEKEQTLL 331
Db  993 -----RKLQAALINRKELLQRVSRLEEEELANLKDESKKEIPLSETERGEVEEDKENKEY 1046

Qy  332 QFQKSKMACQ---LY-REKVNALQAQVCELQKERDQAYSARDSAQ---REISQSLVEKDS 384
Db 1047 SEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEKLAAEEQFQALVKQMNQTLQDK-- 1104

Qy  385 LRRQVFELTDQVCELRTQLRQLQA----------EPPGVLKQEARTREPCPREKQ 429
Db 1105 -----TNQIDLLQAEISENQAIIQKLITSNTDASDGDSVALVKETVVISPPCTGSSE 1156

Qy  430 RLVRMHAICPRDDSDCSLVSSTESQLLSDL-SATSSRELV-----DSFRSSSPAPPSQQS 483
Db 1157 HWK------PELEEKILALEKEKEQLQKKLQEALTSRKAILKKAQEKERHLREELKQQKD 1210
```

```
Qy    9 SALTALDEE---TLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLD-EEVLHSPRLT 64
Db  703 SAFTALSEERDQLLSQVKELSMVTELRAQV-KQLEMNLAEAERQRRLDYESQTAHDNLLT 761

Qy   65 NSAMRAGHLLDLLKTRGKNGAIAFLES------LKFHNPDVYTLVTGLQPDVDFSNFSGL 118
Db  762 EQI----HSLS-IEAKSKDVKIEVLQNELDDVQLQFSEQS--TLIRSLQSQLQ------- 807

Qy  119 METSKLTECLAGA-----IGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETR--AEGLH 171
Db  808 ---NKESEVLEGAERVRHISSKVEELSQALSQKELEITKMDQL----LLEKKRDVETLQ 859

Qy  172 QLEADHSRMKREVSAHFHE-VLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLL- 229
Db  860 QTIEEKDQQVTEISFSMTEKMVQLNEEKFSLGV----------------EIKTLKEQLNLLS 905

Qy  230 ------KQELQRANMVSS-CELELEQSLRTASDQESGDEELNRLKEENEKLRSLTFSLA 282
Db  906 RAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKEELQHEFDLLKKENEQRK------- 958

Qy  283 EKDILEQSLDEARGSRQELVERIHSLRERAV----------AAERQREQYWEEKEQTLL 331
Db  959 -----RKLQAALINRKELLQRVSRLEEEELANLKDESKKEIPLSETERGEVEEDKENKEY 1012

Qy  332 QFQKSKMACQ---LY-REKVNALQAQVCELQKERDQAYSARDSAQ---REISQSLVEKDS 384
Db 1013 SEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEKLAAEEQFQALVKQMNQTLQDK-- 1070

Qy  385 LRRQVFELTDQVCELRTQLRQLQA----------EPPGVLKQEARTREPCPREKQ 429
Db 1071 -----TNQIDLLQAEISENQAIIQKLITSNTDASDGDSVALVKETVVISPPCTGSSE 1122

Qy  430 RLVRMHAICPRDDSDCSLVSSTESQLLSDL-SATSSRELV-----DSFRSSSPAPPSQQS 483
Db 1123 HWK------PELEEKILALEKEKEQLQKKLQEALTSRKAILKKAQEKERHLREELKQQKD 1176

Qy  484 LYKRVAEDFGEEPWSFSSC------LEIPEGDPGALPGAK-----------AGDPHLDYELLDT 530
Db 1177 DYNRLQEQFDEQSKENENIGDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKA 1236

Qy  531 AD----LPQLESSLQPVSP--------GRLDVSESGVLMRRRPARRI----------LS 567
Db 1237 TEQHHTQPVLESNLCPDWPSHSEDASALQGGTSVAQIKAQLKEIEAEKVELELKVSSTTS 1296

Qy  568 QVTMLAFQGDALLEQISVIGGNLTGIFTIHRVTPGSAADQMALRPGTQIVMDYEASEPLF 627
Db 1297 ELTKKSEEVFQLQEQI-------KTVSHEAEVHAESL 1336

Qy  628 KAVLEDTTLEEAVGL--LRRVDGFCCLSVKVNDGYKRLLQDLEAKVATSGDSFYIRVNL 685
Db 1337 QQKLESSQQIA-GLEHLR-------ELQPKLDELQKLISKKEEDVS------YLSGQL 1381

Qy  686 AMEGRAKGELQVHCNEVLHVTDTMFQGCGCWHAHRVNSYTMKDTAAHGTIPNVSRAQQQL 745
Db 1382 SEKEAALTKIQ--TEIIEQEDLI----------KALHTQLEMQAKEHDER 1419

Qy  746 IALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSSTCF 805
Db 1420 IKQLQ--VELCEMKQKPEIGEESRAKQQIQRKIQAALI------------SRKEALKE- 1464

Qy  806 WAESCLTLVPYTLVWPHRPARPRPVLLVPRAVCKILSEKLCLLQGFKKCLAEYLSQEEYE 865
Db 1465 ----------NKSLQEELSLARGTIERLTKSLADVESQ 1492

Qy  866 AWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQ-LDSVCTLHRMDIFPIVH 924
Db 1493 VSAQN---KEKDTVLGRLALLQEERDKLITEMDRSLLENQSLSSSCESLKALEGL--- 1545

Qy  925 VSVNEKMAKKLK------------LGTSE-------EQLLEAA 953
Db 1546 TEDKEKLVKEIESLKSSKIAESTEWQEKHKELQKEYEILLQSYENVSNEAERIQHVVEAV 1605
```

```
Qy   484  LYKRVAEDFGEEPWSFSSC------------AGDPHLDYELLDT  530
          ||| |: ||  :||
Db  1211  DYNRLQEQFDEQSKENENIGDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKA 1270

Qy   531  AD----LPQLESSLQPVSP---------GRLDVSESGVLMRRRPARRI------LS  567
          :: || ||
Db  1271  TEQHHTQPVLESNLCPDWPSHSEDASALQGGTSVAQIKAQLKEIEAEKVELELKVSSTTS 1330

Qy   568  QVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLF  627
          ||| |:|||
Db  1331  ELTKKSEEVFQLQEQI----NKQGLEIESL---------KTVSHEAEVHAESL       1370

Qy   628  KAVLEDTTLEEAVGL-LRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNL   685
          || :  || |||
Db  1371  QQKLESSLQIA-GLEHLR------ELQPKLDELQKLISKKEEDVS-----YLSGQL     1415

Qy   686  AMEGRAKGELQVHCNEVLHVTDTMFQCGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQL   745
          : | :  ||  :| |
Db  1416  SEKEAALTKIQ---TEIIEQEDLI-----------KALHTQLEMQAKEHDER        1453

Qy   746  IALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSSTCF  805
          |:
Db  1454  IKQLQ--VELCEMKQKPEEIGESRAKQQIQRKLQAALI------SRKEALKE--       1498

Qy   806  WAESCLTLVPYTLVWPHRPARPPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYE   865
          : : || |: |::
Db  1499  ---------------------NKSLQEELSLARGTIERLTKSLADVESQ            1526

Qy   866  AWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQ-LDSVCTLHRMDIFPIVIH  924
          :: : :||
Db  1527  VSAQN---KEKDTVLGRLALLQEERDKLITEMDRSLLENQSLSSSCESLKLALEGL--   1579

Qy   925  VSVNEKMAKKLK-----------KGLQR------LGTSE--------EQLLEAA       953
          ::  ::|
Db  1580  TEDKEKLVKEITESLKSSKIAESTEWQEKHKELQKEYEILLQSYENVSNEAAERIQHVVEAV 1639

Qy   954  RQEEGDLDRAPCLYSSL----------APDGWSDLDGLLSCVRQAIADEQKKVVWTE    1000
          |||: |:           ::: :: : :||:||:||
Db  1640  RQEKQE-------LYGKLRSTEANKKETEKQLQEAEQEMEEMEKMRKFAKSKQQKILELE  1693

Qy  1001  Q 1001
          |
Db  1694  E 1694
```

RESULT 13

JE0366

tight junction protein, ZO-2 - chicken

C;Species: Gallus gallus (chicken)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jul-2000
C;Accession: JE0366
R;Collins, J.R.; Rizzolo, L.J.
Biochem. Biophys. Res. Commun. 252, 617-622, 1998
A;Title: Protein-binding domains of the tight junction protein, ZO-2, are highly conserv
A;Reference number: JE0366; MUID:99057550
A;Accession: JE0366
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1163 <COL>
A;Cross-references: GB:AF085184; NID:g3820579; PIDN:AAC95469.1; PID:g3820580

Query Match   4.7%; Score 242; DB 2; Length 1163;
Best Local Similarity 22.2%; Pred. No. 4.5e-05;
Matches 147; Conservative 102; Mismatches 239; Indels 174; Gaps 30;

```
Qy   376  SQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMH  435
          :: |:|| ||  ::
Db   333  ARKLIEKSRGKLQLVLRDR------------------KQTLLNIP                 360

Qy   436  AICPRDDSDCSL--VSSTESQLL-----SDLSATSSRELVDSFRSSSPAPPSQQS       483
          :  :||
Db   361  SL--NDSDSEMDDISEIESNRSFSPQDDRLHHSDLDLHSSNEKLKEKTNAKDDPSNRMS   417
```

```
Qy   484  LYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLTADLPQLESSLQPV   543
          ||  :       ::| :|
Db   418  ------RMGAMPTPFKSTGDI------ATTAVTVVDANKEPKYQDDVAVPQPKAVTRTI   464

Qy   544  -SPGRLDVSESGVLMRRRPARRILSQVTMLAF-QGDALLEQISVIGGNLTGIFIHRVTPG  601
          ::: | ||  ::  |||
Db   465  LKPSPEDEAIYG------PNTKMVRFKKGDSV--GLRLAGGNDVGIFIAGIQEG        510

Qy   602  SAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRV---DGFCCLSVKVNT  658
          :||| |  ||| | |||
Db   511  TSADQEGLQEGDQILKVNAQD---FRGIVR----EDAVLYLLEIPKGD----TVTILA    557

Qy   659  DGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG-CGCWH  717
          :: :|:   |||||:|   ||:|||  ::|
Db   558  QSKYEVYRDIMA--CGRGDSFFIRSHFECEKESPQSLAFTRGEIFRVDTLYDGKLGNWL   615

Qy   718  AHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLVRIVSMD  777
          |: ||| :||| |      ||:|||  ::
Db   616  AVRIGNELEK------GLIPNSRAEQ--MASVQNAQKD-----------GPS-------D  650

Qy   778  KA-------KASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPPV   830
          || |||  |||
Db   651  RADFWRTRGQRSGVKKNLRKSREDLTAIVSVGTKPFA-------YERVQLREAGFKRPV  702

Qy   831  LLVPRAVGKILSEKLC-----LLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCW  885
          :|      :| :::  :|:  ||
Db   703  VIF-GPIADVAMEKLSTDLPHLYQTAK--------TEPRDAGSEKATGV-----------  742

Qy   886  VTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRL-   942
          |   :|||||| ||||  :| |  :|||:||
Db   743  VRLNTVRQIIEQDKHALLDVTPKAVDLLNYTQWFFPIVVFFNPDSKQGVKTMR--QRLCST 800

Qy   943  -GTSEEQLLEAARQEEGDLDR--APCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWT  999
          |: : ::  || |||: ||:    ||
Db   801  SNKSSRKLYEQANKLKKTCSHLFTATINLNSANDSW------YGSLKDTIQQQGEAVWV   854

Qy  1000  EQ 1001
          ::
Db   855  SE 856
```

RESULT 14

I46236

tight junction protein - dog (fragment)

C;Species: Canis lupus familiaris (dog)
C;Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 16-Dec-1998
C;Accession: I46236
R;Jesaitis, L.A.; Goodenough, D.A.
J. Cell Biol. 124, 949-961, 1994
A;Title: Molecular characterization and tissue distribution of ZO-2, a tight junction
A;Reference number: A54475; MUID:94179414
A;Accession: I46236
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-775 <JES>
A;Cross-references: GB:L27152; NID:g463047; PID:g507892
C;Genetics:
A;Gene: ZO-2
C;Superfamily: guanylate kinase homology; GLGF domain homology
F;101-171/Domain: GLGF domain homology <GLG3>
F;316-464/Domain: guanylate kinase homology <GKI>

Query Match   4.6%; Score 236.5; DB 2; Length 775;
Best Local Similarity 23.5%; Pred. No. 5.1e-05;
Matches 133; Conservative 86; Mismatches 231; Indels 115; Gaps 26;

```
Qy   454  QLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGAL  513
          |     :: :|:|   :||
Db     6  QQYSDYDYHSSNEKL-----KERPNSREDMQNRWSR-MGATPTPFKSM-----GDIASV   53

Qy   514  PGAK-AGDPHLDYELLTADLPQLESSLQPVSPGRLDVSESGVLMRRRPARRIL--SQVT  570
```

```
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleo
F;84-764/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;552-565/Region: actin binding #status predicted
F;626-640/Region: actin binding #status predicted
F;837-1936/Domain: coiled coil #status predicted <COI>
F;837-1277/Region: S2
F;1278-1959/Region: light meromyosin
F;1937-1959/Domain: carboxyl-terminal <CBT>
F;125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;694,704/Active site: Cys #status predicted

Query Match             4.6%;  Score 235;  DB 1;  Length 1959;
Best Local Similarity   20.8%;  Pred. No. 0.0002;
Matches  244;  Conservative  170;  Mismatches  385;  Indels  376;  Gaps  58;

Qy      14 LDEETLWEMMESHRH---RIVRCICPS------RLTPYLRQAKVLCQLDEEVLHSPR--  62
              ||:|||  :|   :||:  :|  |:    ||||  ||:
Db     649 LYKEQLAKLMATLRNTNPNFVRCIIPNHEKKAGKLDPHL---VLDQLRCNGVLEGIRIC 704

Qy      63 ------------------------------LTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTL- 102
                                          ||:    ||  |  |:: |:||| |:
Db     705 RQGFPNRVVFQEFRQRYEILTPNAIPKGFM----DGKQACVLMIKALEL-DSNLYRIG 757

Qy     103 ------------------VTGLQPDVDFSNFSGLMETSKLTECLAG--------AIGSLQEELN 140
                                :|:|          :|:
Db     758 QSKVFFRAGVLAHLEERDL---------------KITDVIIGFQACCRGYLARKAFAKRQQQLT 807

Qy     141 QEKGQKEVLLRRCQQLQEHLGLAE---------TRAEGLHQLEADHSRMKREVSAHFHEVLR 193
              ||  |      |::||                ||
Db     808 AMK---VLQRNC---AAYLKLRNWQWWRLFTKVKPLLQV---SRQEEMMAKEEELIK 856

Qy     194 LKDEMLSLSLHYSNALQEKELAASRCRSLQELYLLKQELQRANMVSSCELEQEQSLR-  252
              :|:|                |:||:|       ||||             ||
Db     857 VKEKQLAA---ENRLSEME--------TFQAQLMAEKMQLQE-QLQAEAELCAEAEIRA 904

Qy     253 --TASDQESGDEEL-----------NRLKEENEKLRSLTFSLAEK-------DILEQSLDEARGSR 298
                |   |   ||               ||:|||:|||||   |:||
Db     905 RLTAKKQEL--EEICHDLEARVEEEERCQHLQ---AEKKKMQQNIQELEEQLEEEESAR 959

Qy     299 QEL------VERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREK----- 346
                |       |:||| |:|  |||      |   ||||:|||
Db     960 QKLQLEKVTTEAKLKKLEEDVIVLEDQNLKL--AKEKKLLEDRMSEFTTNLTEEEEKSKS 1017

Qy     347 --VNALQAQVCELQKERDQAYSARDSAQR---EISQSLVEKDS--LRRQVFELTDQVC 397
                :|    |  ||||::|    ::|            ::||:|    ||||:|| ||
Db    1018 LAKLKNKHEAMTDLEER-----LRREEKQRQELEKTRRKLEGDSSDLHDQIAELQAQIA 1072

Qy     398 ELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLS 457
              ||:||                |:||  ||:    ||| || | |:::::
Db    1073 ELKIQLSKKEEELQAAL--ARVEEEAAQKNMALKKIREL------ESQI--1113

Qy     458 DLSATSSSRELVDSFRSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAK 517
                :||                   ||||||
Db    1114 ----TELQEDLESERASRNKAEKQK---------RDLGEELEALKTELEDTLDSTAAQELR 1162

Qy     518 AGDPH----LDYELLTDADL--PQLESSLQPVSPGRLDVSESGVLMRRPA----- 562
              : |                    |||:||
Db    1163 SKREQEVTVLKKTLEDEAKTHEAQIQEMRQKHSQAEELAEQLEQTKRVKANLEKAKQAL 1222

Qy     563 ---RRILSQVTMLAFQG---------DALLEQISVIGGNLTGIFIHRVTPGSAADQMA 608
                   |      |            ||
Db    1223 ESERAELSNEVKVLLQGKGDAEHKRKKVDAQLQELQV----------KFTEGE-- 1265

Qy     609 LRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTD--GYRKLLQ 666
              |||                  ||||||:
Db    1266 ---------RVKTELAERVNKLQVE--LDNVTGLLNQSDS---KSIKLAKDFSALESQLQ 1311

Qy     667 DLEAKV----------------ATSGDSFYIRVNLAMEGRAKGELQ---VHCNEVLHVTD 707
              |||:|                 ||
Db    1312 DTQELIQEETRIKLSFSTKLKQTEDERNALKEQLEEEEAKRNLEKQISVLQQQAVEARK 1371
```

RESULT 15
A33977
myosin heavy chain, nonmuscle - chicken
N;Contains: myosin ATPase (EC 3.6.1.32)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
C;Accession: A33977; S06116; A43422
R;Shohet, R.V.; Conti, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989
A;Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my
A;Reference number: A33977; MUID:90046668
A;Accession: A33977
A;Molecule type: mRNA
A;Residues: 1-1959 <SHO>
A;Cross-references: GB:M26510; NID:g212382; PIDN:AAA48974.1; PID:g212383
R;Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.
Eur. J. Biochem. 184, 611-616, 1989
A;Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in
s.
A;Reference number: S06116; MUID:90033648
A;Accession: S06116
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 716-1008 <KAT>
A;Cross-references: GB:X17589
A;Note: this translation is not annotated in GenBank entry GGMHCFMHA, release 114
R;Hodge, T.P.; Cross, R.; Kendrick-Jones, J.
J. Cell Biol. 118, 1085-1095, 1992
A;Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate
A;Reference number: A43422; MUID:92381096
A;Accession: A43422
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1900-1959 <HOD>
A;Experimental source: brush border
A;Note: sequence extracted from NCBI backbone (NCBIP:111947)
C;Superfamily: myosin heavy chain; myosin motor domain homology

```
Db      54 VGTENSKEPRYQEE--------------PPAPQKAAPR------TFLRPSPEDEAIYGPNTK  96

Qy     571 MLAF-QGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKA 629

Qy      97 MVRFKKGDSV--GLRLAGGNDVGIFVAGIQEGTSAEQEGLQEGDQILKVNTQDFRGL--- 151

Qy     630 VLEDTTLEEAVGLLRRVDG-FCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAME 688

Db     152 VREDAVLY----LLEIPKGEMVTILAQSRADVIRDIL-----ACGRGDSFFIRSHFECE 201

Qy     689 GRAKGELQVHCNEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIA 747

Db     202 KETPQSLAFSRGEVFRVVDTLYDGKLGHWLAVRIGNELEK-----GLIPNKSRAEQ--MA 254

Qy     748 LIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSSTCFWA 807

Db     255 SVQN------AQRDNAGDRADFEWRM---RGQRSGMKKNLRKSREDLTAAVSVSTKFPA 303

Qy     808 ESCLTLVPYTLVWPHRPARPRPVLL---VPRAVGKILSEKLCLLQGFKKCLAEYLSQEE 863

Db     304 ------YERVLLREAGEKRPVVLFGPIADIALEKLANELPDLFQTAK-------TEP 347

Qy     864 YEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVI 923

Db     348 KDAGSEKSSGV-------VRLNTVRQIIEQDKHAILDVTPKAVDLLNYTQWFPIVI 396

Qy     924 HVSVNEKMAKKLKKGLQRLGTSEEQLLAAARQEEGDLDRAPCLY-------SSLAPDGWS 976

Db     397 FFNPDSRQGVKTMR--QRLNPTSNKSSRKLYDQANKLKKT-CAHLFTATINLNSANDSW-452

Qy     977 DLDGLLSCVRQAIADEQKKVVWTEQ 1001

Db     453 -----FGSLKDTIQHQQGEAVWVSE 472
```

```
Qy    708 TMFQGCGCWH-AHRVNSYTMKD---------TAAHGTI-PNYSRAQQQLIALIQDMT- 753
              : | :| | |      ||              ||:  :  :| ||:| : |:
Db   1372 KMDDGLGCLEIAEEAKKKLQKDLESLTQRYEEKIAAYDKLEKTKRLQQELDDIAVDLDH 1431

Qy    754 QQCTVTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTL 813
              |:·||:   :| ::::|·:| ::              |        |:|
Db   1432 QRQTVSNLEKK--QKKFDQLLAEEKNISAKYAEERDRAEAEAREKE--------TKALSL 1481

Qy    814 VPYTLVWPHRPARPRPVLLVPRA--------------------VGKILSEKLCLLQGF 851
              ||   :|                              ||| | |  |:
Db   1482 --------ARALEEAIEQKAELERVNKQFRTEMEDLMSSKDDVGKSVHE----LEKA 1526

Qy    852 KKCLAEYLSQ-----EEYEAWSQRGDIIQEGEVSGGRCWVTRHAV----------ESL 894
              |:·|:|::     ||·|       |·:|:|:·|:·|:            |
Db   1527 KRALEQQVEEMKTQLEELE------DELQATEDAKLRLEVNQQAMKAQFDRDLLGRDEQN 1580

Qy    895 MEKNTHAL-----LDVQLD------SVCTLHR----MDIFPIVIHVSVNEK---------- 930
              ||       ::|:|:       |:         :|| :  |:
Db   1581 EEKRKQLIRQVREMEVELEDERKQRSIAVAARKKLELDLKDLESHIDTANKNRDEAIKHV 1640

Qy    931 -----MAKKLKKGLQRLGTSEEQLLEAARQEEGDL 960
               :  |:  ||:||:| ||:| |
Db   1641 RKLQAQMKDYMRELEDTRTSREEILAQAKENEKKL 1675
```

Search completed: June 13, 2002, 09:24:56
Job time: 154 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:19:57 ; Search time 39.24 Seconds
                                          (without alignments)
                                          2841.949 Million cell updates/sec

Title: US-09-767-215-2
Perfect score: 5149
Sequence: 1 MGELCRRDSALTALDEETLW........VRQAIADEQKKVVWTEQSPR 1004

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_032802:*
    1: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1980.DAT:*
    2: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1981.DAT:*
    3: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1982.DAT:*
    4: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1983.DAT:*
    5: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1984.DAT:*
    6: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1985.DAT:*
    7: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1986.DAT:*
    8: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1987.DAT:*
    9: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1988.DAT:*
   10: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1989.DAT:*
   11: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1990.DAT:*
   12: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1991.DAT:*
   13: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1992.DAT:*
   14: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1993.DAT:*
   15: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1994.DAT:*
   16: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1995.DAT:*
   17: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1996.DAT:*
   18: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:*
   19: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:*
   20: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:*
   21: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:*
   22: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5149	100.0	1004	22	AAE07164	Human caspase recr
2	5034	97.8	1139	22	AAE07165	Human predicted ca
3	1239.5	24.1	1147	22	AAU01207	Human caspase recr
4	887.5	17.2	1032	22	AAU01206	Human caspase recr
5	566.5	11.0	746	22	AAM93822	Human polypeptide,
6	512.5	10.0	536	22	AAU01204	Rat caspase recrui
7	493	9.6	536	22	AAU01205	Human caspase recr
8	396.5	7.7	366	22	AAB95617	Human protein sequ
9	309	6.0	350	22	AAM38934	Human polypeptide
10	282	5.5	1445	22	AAB64619	Drosophila melanog
11	262	5.1	2101	15	AAR47173	Sequence of the in
12	262	5.1	2101	22	AAG65799	Nuclear mitotic ap
13	262	5.1	2207	21	AAU32041	Novel human secret
14	259	5.0	2101	21	AAY49936	Human NuMA protein
15	258	5.0	1916	22	ABB62423	Drosophila melanog
16	258	5.0	1916	22	ABB66089	Drosophila melanog
17	254.5	4.9	2246	22	ABG05850	Novel human diagno
18	252	4.9	2217	22	AAU32040	Novel human secret
19	252	4.9	2192	18	AAW21732	LexA/NuMA fusion p
20	252	4.9	2272	18	AAW21731	GAL4/HA/NuMA fusio
21	248.5	4.8	2442	21	AAY77575	Human cytoskeletal
22	246.5	4.7	2115	21	AAY49937	Human NuMA protein
23	240.5	4.7	1960	22	AAM78854	Human protein SEQ
24	238	4.6	3248	17	AAR99795	Kinetochore protei
25	237	4.6	2482	16	AAR72826	Human mitosin. Ho
26	237	4.6	2482	19	AAW23996	Human mitosin amin
27	236.5	4.6	2096	21	AAB41592	Human ORFX ORF1356
28	236.5	4.6	2139	22	AAB47278	PN7771. Homo sapi
29	235	4.6	612	22	AAB95546	Human protein sequ
30	233.5	4.5	1456	22	ABB58673	Drosophila melanog
31	233	4.5	1963	22	AAM79838	Human protein SEQ
32	232.5	4.5	1037	22	ABG23862	Novel human diagno
33	230	4.5	931	22	ABG20279	Novel human diagno
34	229.5	4.5	2415	22	ABG22366	Novel human diagno
35	228	4.4	674	19	AAW72748	Human P-dlg protei
36	228	4.4	1717	22	ABG20672	Novel human diagno
37	226.5	4.4	2048	22	AAM40027	Human polypeptide
38	223.5	4.3	2400	22	ABG20278	Novel human diagno
39	220.5	4.3	1374	22	AAB69070	Human male enhance
40	217.5	4.2	1975	22	ABB62094	Drosophila melanog
41	217	4.2	1972	17	AAW00024	Smooth muscle myos
42	217	4.2	2143	22	ABG01716	Novel human diagno
43	216.5	4.2	1372	19	AAW56473	Protein with Rho p
44	212.5	4.1	1489	22	ABB59948	Drosophila melanog
45	212.5	4.1	1988	22	AAM40999	Human polypeptide

ALIGNMENTS

RESULT 1
AAE07164
ID   AAE07164 standard; Protein; 1004 AA.
XX
AC   AAE07164;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human caspase recruitment domain-14 (CARD-14).
XX
KW   Human; caspase recruitment domain-14; CARD-14; chromosome 17;
KW   nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
KW   autoimmune disorder; systemic lupus erythematosus; neurological disorder;
KW   Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
KW   haematological disorder; myelodysplastic syndrome; myocardial infarction;
KW   stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
KW   cell signalling disorder; cytostatic; immunosuppressive; nootropic;
KW   neuroprotective; antiviral; antibacterial.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Modified-site   6.9
FT   /note= "cAMP- and cGMP-dependent protein kinase
FT   phosphorylation site"
FT   Domain   10..116
FT   /label= CARD_domain
FT   Modified-site   12..15
FT   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   18..21
FT   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   25..27
FT   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   60..62

```
FT   Modified-site   91..93
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   114..117
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   117..122
FT                   /note="N-glycosylation site"
FT   Modified-site   121..123
FT                   /note="N-myristoylation site"
FT   Modified-site   126..128
FT                   /note="Protein kinase C phosphorylation site"
FT   Domain          130..420
FT                   /label= Coiled_Coil_domain
FT   Modified-site   134..137
FT                   /note="N-myristoylation site"
FT   Modified-site   161..166
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   165..168
FT                   /note="N-myristoylation site"
FT   Modified-site   220..227
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   221..224
FT                   /note="Tyrosine kinase phosphorylation site"
FT   Domain          239..325
FT                   /label= k-Box_domain
FT   Modified-site   240..243
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   250..252
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   253..256
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   259..262
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   280..283
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   290..293
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   297..300
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   307..309
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   307..310
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   359..365
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   366..368
FT                   /note="Tyrosine kinase phosphorylation site"
FT   Modified-site   366..369
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   378..381
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   384..386
FT                   /note="Casein kinase II phosphorylation site"
FT   Region          385..406
FT                   /note="Leucine zipper pattern"
FT   Modified-site   449..452
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   463..466
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   463..465
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   470..472
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   501..504
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   511..516
FT                   /note="Casein kinase II phosphorylation site"
FT   Domain          568..660
FT                   /label= PDZ_domain
FT   Modified-site   587..592
FT                   /note="N-myristoylation site"
FT   Modified-site   589..592
FT                   /note="N-glycosylation site"
FT   Modified-site   602..605
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   634..637
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   653..655
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   674..677
FT                   /note="Casein kinase II phosphorylation site"
FT   Domain          676..745
FT                   /label= SH3_domain
FT   Modified-site   714..719
FT                   /note="N-myristoylation site"
FT   Modified-site   725..727
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   725..728
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   733..738
FT                   /note="N-myristoylation site"
FT   Modified-site   737..740
FT                   /note="N-glycosylation site"
FT   Modified-site   759..761
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   760..763
FT                   /note="cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Peptide         785..793
FT                   /note="Peroxisomal targeting signal"
FT   Modified-site   796..799
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   800..805
FT                   /note="N-myristoylation site"
FT   Domain          826..1004
FT                   /label= Guanylate_kinase_domain
FT   Modified-site   842..844
FT                   /note="Protein kinase C phosphorylation site"
FT   Modified-site   860..863
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   868..870
FT                   /note="Protein kinase C phosphorylation site"
FT   Region          870..872
FT                   /note="RGD cell attachment sequence"
FT   Modified-site   893..896
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   926..929
FT                   /note="Casein kinase II phosphorylation site"
FT   Peptide         941..949
FT                   /note="Peroxisomal targeting signal"
FT   Modified-site   944..947
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   976..979
FT                   /note="Casein kinase II phosphorylation site"
FT   Modified-site   980..985
FT                   /note="N-myristoylation site"
FT   Modified-site   1002..1004
FT                   /note="Protein kinase C phosphorylation site"
XX
PN   WO20015065-A2.
XX
PD   16-AUG-2001.
XX
PF   22-JAN-2001; 2001WO-US02087.
XX
PR   09-FEB-2000; 2000US-0181159.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-497073/54.
DR   N-PSDB; AAD13447.
XX
PT   An isolated caspase recruitment domain polypeptide useful for
PT   regulating growth and cell death and useful for the treatment of cancer
```

PT   -
XX
PS   Claim 1; Fig 1A-1E; 109pp; English.
XX
CC   The present sequence is human caspase recruitment domain-14 (CARD-14).
CC   The CARD-14 gene is located on chromosome 17. The CARD-14 is used for
CC   the detection of modulators that modulates the ability of CARD-14 to
CC   bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or activation
CC   of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for regulating
CC   growth and cell death and useful for the treatment of cancer. It is
CC   also useful for the treatment of autoimmune disorders (e.g., systemic
CC   lupus erythematosus), neurological disorders e.g., Alzheimer's and
CC   Parkinson's disease, inflammatory disorders, haematological disorders
CC   (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
CC   strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
CC   cell signalling disorders and certain viral and bacterial infections.
XX
SQ   Sequence 1004 AA;

Query Match   100.0%; Score 5149; DB 22; Length 1004;
Best Local Similarity   100.0%;
Matches 1004; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MGELCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 mgelcrrdsaltaldeetlwemmeshrhrivrcicpsrltpylrqakvlcqldeeevlhs 60

Qy   61 PRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLME 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 prltnsamraghlldllktrgkngaiaflesplkfhnpdvytlvtglqpdvdfsnfsglme 120

Qy   121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQOLQEHLGLAETRAEGLHQLEADHSRM 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 tsklteclagaigslqeelnqekgqkevllrrcqqlqehlglaetraeglhqleadhsrm 180

Qy   181 KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 krevsahfhevlrlkdemlslslhysnalqekelaasrcrslqeelyllkqelqranmvs 240

Qy   241 SCELELQEOSLRTASDQESGDEELNRLKEENEKLRSLTFSLAEKDILEQSLDEARGSRQE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 scelelqeqslrtasdqesgdeelnrlkeeneklrsltfslaekdileqsldeargsrqe 300

Qy   301 LVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKE 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 lverihslreravaaerqreqyweekeqtllqfqkskmacqlyrekvnalqaqvcelqke 360

Qy   361 RDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEART 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 rdqaysardsaqreisqslvekdslrrqvfeltdqvcelrtqlrqlqaeppgvlkqeart 420

Qy   421 REPCPREKQRLVRMHAICPRDDSDCCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPS 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 repcprekqrlvrmhaicprddsdccslvsstesqllsdlsatssrelvdsfrssspapps 480

Qy   481 QQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSL 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 qqslykrvaedfgeepwsfsscleipegdpgalpgakagdphldyelldtadlpqlessl 540

Qy   541 QPVSPGRLDVSESGVLMRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 qpvspgrldvsesgvlmrrparrilsqvtmlafqgdalleqisviggnltgifihrvtp 600

Qy   601 GSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 gsaadqmalrpgtqivmvdyeaseplfkavledttleeavgllrrvdgfcclsvkvntdg 660

Qy   661 YKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQGCGCWHAHR 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ykrllqdleakvatsgdsfyirvnlamegrakgelqvhcnevlhvtdtmfqgcgcwhahr 720

Qy   721 VNSYTMKDTAAHGTIPNYSRAQQOLIALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDKAK 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 vnsytmkdtaahgtipnysraqqolialiqdmtqqctvtrkpssggpqklvrivsmdkak 780

Qy   781 ASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 asplrlsfdrgqldpsrmegsstcfwaescltlvpytlvwphrparprpvllvpravgki 840

Qy   841 LSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTH 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 lseklcllqgfkkclaeylsqeeyeawsqrgdiiqegevsggrcwvtrhaveslmeknth 900

Qy   901 ALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQORLGTSEEQLLEAARQEEGDL 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 alldvqldsvctlhrmdifpivihvsvnekmakklkkglqrlgtseeqlleaarqeegdl 960

Qy   961 DRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTEQSPR 1004
          ||||||||||||||||||||||||||||||||||||||||||||
Db   961 drapclysslapdgwsdldgllscvrqaiadeqkkvvwteqspr 1004

RESULT   2
AAE07165
ID   AAE07165 standard; Protein; 1139 AA.
XX
AC   AAE07165;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human predicted caspase recruitment domain-14 (CARD-14).
XX
KW   Human; caspase recruitment domain-14; CARD-14; chromosome 17;
KW   nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
KW   autoimmune disorder; systemic lupus erythematosus; neurological disorder;
KW   Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
KW   haematological disorder; myelodysplastic syndrome; myocardial infarction;
KW   stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
KW   cell signalling disorder; cytostatic; immunosuppressive; nootropic;
KW   neuroprotective; antiviral; antibacterial.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 700
FT   /note= "Encoded by TGG"
XX
PN   WO200159065-A2.
XX
PD   16-AUG-2001.
XX
PF   22-JAN-2001; 2001WO-US02087.
XX
PR   09-FEB-2000; 2000US-0181159.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-497073/54.
DR   N-PSDB; AAD13448.
XX
PT   An isolated caspase recruitment domain polypeptide useful for
PT   regulating growth and cell death and useful for the treatment of cancer
PT   -
XX
PS   Disclosure; Fig 2A-2C; 109pp; English.
XX
CC   The present sequence is predicted human caspase recruitment domain-14
CC   (CARD-14). The CARD-14 gene is located on chromosome 17. The CARD-14 is
CC   used for the detection of modulators that modulates the ability of
CC   CARD-14 to bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or
CC   activation of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for
CC   regulating growth and cell death and useful for the treatment of cancer.
CC   It is also useful for the treatment of autoimmune disorders (e.g.,

```
CC    systemic lupus erythematosus), neurological disorders e.g., Alzheimer's
CC    and Parkinson's disease, inflammatory disorders), haematological disorders
CC    (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
CC    strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
CC    cell signalling disorders and certain viral and bacterial infections.
XX
SQ    Sequence    1139 AA;

Query Match             97.8%;  Score 5034;  DB 22;  Length 1139;
Best Local Similarity   95.1%;  Pred. NO. 0;
Matches  995;  Conservative    0;  Mismatches    0;  Indels   1;  Gaps   3;

Qy          1 MGELCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHS  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 mgelcrrdsaltaldeetlwemmeshrhrivrcicpsrltpylrqakvlcqldeeevlhs  60

Qy         61 PRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFS----  116
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 prltnsamraghllldllktrgkngaiaflesllkfhnpdvytlvtglqpdvdfsnfsgess  120

Qy        117 --------------GLMETSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQE  158
                             ||||||||||||||||||||||||||||||||||||||
Db        121 dfdglagtsrnlrilvtpglmetsklteclagaigslqeelnqekgqkevllrrcqqlqe  180

Qy        159 HLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASR  218
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 hlglaetraeglhqleadhsrmkrevsahfhevlrlkdemlslslhysnalqekelaasr  240

Qy        219 CRSLQEELYLLKQELQRANMVSSCELELQQSLRTASDQESGDEELNRLKEENEKLRSLT  278
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 crslqeelyllkqelqranmvsscelelqeqslrtasdqesgdeelnrlkeeneklrslt  300

Qy        279 FSLAEKDILEQSLDEARGSRQELVERIHSLRERAVAAERQRE----------------  320
              |||||||||||||||||||||||||||||||||||||||||
Db        301 fslaekdileqsldeargsrqelverihslreravaaerqreqarpsellsftvhvshsv  360

Qy        321 QYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYSARDSAQREISQSLV  380
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 qyweekeqtllqfqkskmacqlyrekvnalqaqvcelqkerdqaysardsaqreisqslv  420

Qy        381 EKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPR  440
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ekdslrrqvfeltdqvcelrtqlrqlqaeppgvlkqeartreppprekqrlvrmhaicpr  480

Qy        441 DDSDCSLVSSTESQLLSDLSATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFS  500
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ddsdcslvsstesqllsdlsatssrelvdsfrssspappsqqslykrvaedfgeepwsfs  540

Qy        501 SCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSE-------  552
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 scleipegdpgalpgakagdphldyelldtadlpqlesslqpvspgrldvsesesagqrlp  600

Qy        553 --SGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFTIHRVTPGSAADQMALR  610
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 acsgvlmrrrparrilsqvtmlafgqdalleqisviggnltgifihrvtpgsaadqmalr  660

Qy        611 PGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEA  670
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 pgtqivmvdyeaseplfkavledttleeavgllrrvdgfcclsvkvntdgykrllqdlea  720

Qy        671 KVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQGCGCWHAHRVNSYTMKDTA  730
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 kvatsgdsfyirvnlamegrakgelqvhcnevlhvtdtmfqgcgcwhahrvnsytmkdta  780

Qy        731 AHGTIPNYSRAQQQLIALIQDMTQQCTVRKPSSGGPQKLVRIVSMDKAKASPLRLSFDR  790
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ahgtipnysraqqqlialiqdmtqqctvtrkpssggpqklvrivsmdkakaspirlsfdr  840

Qy        791 GQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQG  850
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 gqldpsrmegsstcfwaescltlvpytlvwphrparprpvllvpravgkilseklcllqg  900

Qy        851 FKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSV  910
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 fkkclaeylsqeeyeawsqrgdiiqegevsggrcwvtrhavesimeknthalldvqldsv  960

Qy        911 CTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDLRAPCLYSSL  970
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ctlhrmdifpivihvsvnekmakklkkglqrlgtseeqlleaarqeegdlrapclyssl  1020

Qy        971 APDGWSDLDGLLSCVRQAIADEQKKV  996
              |||||||||||||||||||||||||
Db       1021 apdgwsdldgllscvrqaiadeqkkv  1046

RESULT   3
AAU01207
ID   AAU01207 standard; Protein; 1147 AA.
XX
AC   AAU01207;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Human caspase recruitment domain, CARD-11 polypeptide.
XX
KW   Human; caspase recruitment domain; CARD-11; Bcl-10; NF-kappaB;
KW   apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW   inflammatory disorder; viral infection; stress-related response.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          6..112
FT                   /note= "CARD domain"
FT   Modified-site   7..9
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   7..10
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   100..102
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   100..103
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   105..107
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   106..109
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Domain          130..431
FT                   /note= "Coiled coil domain"
FT   Modified-site   162..165
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   168..171
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   175..183
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Modified-site   182..185
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   189..195
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Modified-site   241..244
FT                   /note= "N-glycosylation site"
FT   Modified-site   243..245
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   282..285
FT                   /note= "Amidation site"
FT   Modified-site   286..289
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   290..292
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   378..381
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   429..432
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   459..461
```

```
FT   Modified-site   471..474
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   472..475
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   476..479
FT                   /note= "N-glycosylation site"
FT   Modified-site   508..510
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   510..513
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   558..560
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   578..581
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   584..587
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   587..592
FT                   /note= "N-glycosylation site"
FT   Modified-site   634..637
FT                   /note= "N-myristoylation site"
FT   Modified-site
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Domain          635..748
FT                   /note= "PDZ domain"
FT   Domain          635..1147
FT                   /note= "MAGUK domain"
FT   Modified-site   638..641
FT                   /note= "Glycosaminoglycan attachment site"
FT   Modified-site   678..683
FT                   /note= "N-myristoylation site"
FT   Modified-site   687..689
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   692..695
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   698..703
FT                   /note= "N-myristoylation site"
FT   Modified-site   710..715
FT                   /note= "N-myristoylation site"
FT   Modified-site   725..728
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   761..766
FT                   /note= "N-myristoylation site"
FT   Modified-site   764..767
FT                   /note= "Casein kinase II phosphorylation site"
FT   Domain          766..834
FT                   /note= "SH3 domain"
FT   Modified-site   776..779
FT                   /note= "N-glycosylation site"
FT   Modified-site   779..782
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   787..789
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   816..819
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   823..828
FT                   /note= "N-myristoylation site"
FT   Modified-site   847..850
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   853..858
FT                   /note= "N-myristoylation site"
FT   Modified-site   857..859
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   872..875
FT                   /note= "Casein kinase II phosphorylation site"
FT   Domain          882..1147
FT                   /note= "Guanylate kinase (GUK) domain"
FT   Modified-site   897..900
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   917..922
FT                   /note= "N-myristoylation site"
FT   Modified-site   926..929
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   935..937
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   1003..1006
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   1010..1018
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   1050..1055
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Modified-site   1088..1091
FT                   /note= "N-myristoylation site"
FT   Modified-site   1120..1123
FT                   /note= "Casein kinase II phosphorylation site"
FT
XX
PN   WO200140468-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32716.
XX
PR   03-DEC-1999;  99US-0168780.
PR   18-FEB-2000; 2000US-0507533.
PR   25-FEB-2000; 2000US-0513904.
PR   10-OCT-2000; 2000US-0685791.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-367809/38.
DR   N-PSDB; AAS05389.
XX
PT   Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10,
PT   CARD-11, useful as targets for therapy, as immunogens, and in screening
PT   and detection assays -
XX
PS   Claim 9; Fig 14A-14C; 145pp; English.
XX
CC   The present sequence represents novel human caspase recruitment
CC   domain, CARD-11. The polynucleotide encoding this sequence was
CC   isolated from a human T-cell cDNA library. Also described are
CC   novel human sequences for CARD-9 and CARD-10 (AAU01205, AAU01206)
CC   and rat CARD-9 (AAU01204), CARD-9, CARD-10 and CARD-11 interact
CC   with Bcl-10 which is thought to activate nuclear factor (NF)-kappaB
CC   and apoptosis. The sequences of the invention can be used for
CC   treating a disorder associated with abnormal levels of apoptosis by
CC   modulating the expression or activity of CARD-9, CARD-10, or CARD-11.
CC   They can be used for the treatment of hyperproliferative disorders
CC   (e.g. cancer), autoimmune disorders (e.g. systemic lupus erythematosus),
CC   neurological disorders (e.g. Alzheimer's disease), inflammatory disorders
CC   (e.g. Crohn's disease), and viral infection (e.g. HIV). The CARD
CC   polypeptide, polynucleotide and an antibody which selectively binds to
CC   CARD can be used in screening and detection assays (e.g. chromosomal
CC   mapping, tissue typing), predictive medicine (prognostic assays,
CC   monitoring clinical trials, and therapy (treatment and prophylaxis). The
CC   CARD polypeptide may be used to screen for drugs that bind to and/or
CC   modulate it. CARD sequences are potential targets for regulating
CC   inflammation, cancer, NF-kappaB signalling, stress-related response and
CC   apoptosis in human disease. A host cell containing a polynucleotide
CC   encoding CARD can be used to create transgenic animals.
XX
SQ   Sequence  1147 AA;

Query Match             24.1%;  Score 1239.5;  DB 22;  Length 1147;
Best Local Similarity   30.5%;
Matches  354;  Conservative  204;  Mismatches  402;  Indels  201;  Gaps  33;

Qy         15 DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL 74
              :|: |||:  :|| :  ||||||: |:|:|||||:|::|:::||:::|:: ||||
Db         11 eedalwenvecnrhmlsryinpakltpylrqckvideqdedevinapmlpskinragrll 70

Qy         75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPVDFSNFSGLMETSKLTECLAGAIGS 134
              |:| |:|: | |||||| :  |  ||||| :|   ||| ||  ||  ||
```

```
RESULT  4
AAU01206
ID   AAU01206 standard; Protein; 1032 AA.
XX
AC   AAU01206;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Human caspase recruitment domain, CARD-10 polypeptide.
XX
KW   Human; caspase recruitment domain; CARD-10; Bcl-10; NF-kappaB;
KW   apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW   inflammatory disorder; viral infection; stress-related response.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   Modified-site   15..20
FT                   /note= "N-myristoylation site"
FT   Modified-site   18..21
FT   Domain          23..123
FT                   /note= "CARD domain"
FT   Modified-site   68..70
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   76..79
FT                   /note= "N-glycosylation site"
FT   Modified-site   78..80
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   88..91
FT                   /note= "Amidation site"
FT   Modified-site   112..115
FT   Modified-site   113..118
FT                   /note= "Casein kinase II phosphorylation site"
FT                   /note= "N-myristoylation site"
FT   Domain          147..457
FT                   /note= "Coiled coil domain"
FT   Modified-site   201..207
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Region          230..251
FT                   /note= "Leucine zipper homology region"
FT   Modified-site   242..245
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   293..295
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   293..296
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   309..314
FT                   /note= "N-myristoylation site"
FT   Modified-site   313..315
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   331..334
FT                   /note= "Casein kinase II phosphorylation site"
FT   Domain          366..398
FT                   /note= "Tropomyosin domain"
FT   Modified-site   412..415
FT                   /note= "Casein kinase II phosphorylation site"
FT   Region          426..447
FT                   /note= "Leucine zipper homology region"
FT   Modified-site   438..441
FT                   /note= "Casein kinase II phosphorylation site"
FT   Domain          457..1032
FT                   /note= "MAGUK domain"
FT   Modified-site   472..475
FT                   /note= "N-glycosylation site"
FT   Modified-site   478..481
FT                   /note= "N-myristoylation site"
FT   Modified-site   487..492
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   510..513
FT                   /note= "Casein kinase II phosphorylation site"
```

```
Db    71 dilhtkggrgyvvfleslefyypelykivtgkeptrfstivveeghglthflmnevik 130
Qy   135 LQEELNQEKGQKEVLLRRCQQLQ-EHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLR 193
Db   131 lqgqmkakdlqrcellarlrqledekkqmtltrve-lltfqeryykmkeerdsyndelvk 189
Qy   194 LKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEQSLRT 253
Db   190 vkddnynlamryaqlseeknmavmrsrdlqleldqlkhrlnk--meeckle-rnqslkl 246
Qy   254 ASDQES--GDEELNRLKEENEKLRSLTFSL------------AEKDILEQSLDEAR 295
Db   247 kndienrpkkeqvlelerenemlktknqelqsliqagkrsipdsdkaildilehdrkeal 306
Qy   296 GSRQELVERIHSLRERAVAAEROREQTWEEKEQTLLQFQKSRMACQLYREKVNALQAQVC 355
Db   307 edrqelvnriynlqeearqaeelrdkyleekedlelkcstigkdcemykhrmntvmlqle 366
Qy   356 ELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQ-------VCELRT 401
Db   367 evererdqafhsrdeaqtqysqcliekdkyrkjireleekndemriemvrreacivnles 426
Qy   402 QLRQLQAE------------PPGVLKQ------------EARTREPCPREK------ 428
Db   427 klrrlskdsnnldqslprnlpvtiisqdfgdasprtngqeaddsstseespedskyflpy 486
Qy   429 --QRLVRMHAI-CPRDDSDCSLVSSTESQLL------SDLSATS--SRELVDSFRSSSP 476
Db   487 hppqrrmnlkgiqlqrakspislkrtsdfqakgheeegtdaapsscgslpitnsftkmqp 546
Qy   477 --------------APPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKA-GDPHL 523
Db   547 prsrsslmsitaeppgndsivrrykedaphr-----stve-ednsggfdaldldddshe 600
Qy   524 DY-------------ELLDTADLPQLESSLQPVS--PGRLDVSESGVLMRRRPARR 564
Db   601 rysfgpssihssssshqsegldaydleqvnlmfrkfslerpfrpsvtsvghvrgpgps-- 658
Qy   565 ILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASE 624
Db   659 ---vqhttlngdsltsqitlliggnargsfvhsvkpgslaekaglreghqlllllegcirg 714
Qy   625 PLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATSGDSFYIRVN 684
Db   715 erqsvpldctkeeahwtiqrcsgpvtlhykvnhegyrklvkdmedgliitsgdsfyirln 774
Qy   685 LAMEGRAKG--ELQVHCNEVLHVTDTMFQCGCGWHAHRVNSYTMKDTRAHGTIPNYSRAQQ 743
Db   775 lnissqidactmsikcddvvhvrdtmyqdrhewpcarvdpftdhdldm-gtipsystaqq 833
Qy   744 QLIALIQDMTQQ---------------CTVTRKP----SSGGPQ------------KLV 771
Db   834 lllvklqrlmhrgsreevdgthntlralrntlqpeealstsdprvsprlsrasflfgqll 893
Qy   772 RIVSMDKAK--------------ASPL----RLSFD--------RGQLDPSRMEGSSTCF 805
Db   894 qfvsrsenkykrmnsnervriisgsplgslarssldatklltekqeeldpeselgkn--- 950
Qy   806 WAESCLTLVPYTLVPWHRPARPRPVLLVPRAVGKILSEKLCLLQG--FKKCLAEYLSQE 862
Db   951 -----lslipyslvrafycerrrpvliftptviaktlvgrllnsggamefticksdivtrd 1005
Qy   863 EYEAWSQRCDII--QEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFP 920
Db  1006 eflrrqktetilysreknpnafec-iapanieavaaknkhcllleagigctrdliksniyp 1064
Qy   921 IVIHVSVNEKMAKKLKKGLQRLGTSEEQLLEAARQEEGDLDRAPCLYYSSLAPDGWSDLDG 980
Db  1065 ivlfirvceknikfrkllprpet-eeeflrvcrlkekealpclyatvepdmwgsvee 1123
Qy   981 LLSCVRQAIADEQKKVVWTEQ 1001
Db  1124 llrvvkdkigeeqrktiwvde 1144
```

```
FT   Modified-site   512..514
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   549..552
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   558..560
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   565..570
FT                   /note= "N-myristoylation site"
FT   Modified-site   570..573
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   595..598
FT                   /note= "N-glycosylation site"
FT   Modified-site   603..605
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   638..641
FT                   /note= "Glycosaminoglycan attachment site"
FT   Modified-site   642..644
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   656..661
FT                   /note= "N-myristoylation site"
FT   Modified-site   681..684
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   690..693
FT                   /note= "Casein kinase II phosphorylation site"
FT   Domain          704..772
FT                   /note= "SH3 domain"
FT   Modified-site   712..715
FT                   /note= "N-glycosylation site"
FT   Modified-site   714..717
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   733..739
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Modified-site   748..751
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   754..756
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   754..757
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   761..766
FT                   /note= "N-myristoylation site"
FT   Modified-site   782..784
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   809..814
FT                   /note= "N-myristoylation site"
FT   Domain          830..1032
FT                   /note= "Guanylate kinase (GUK) domain"
FT   Modified-site   830..832
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   868..870
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   869..872
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   882..885
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   893..898
FT                   /note= "N-myristoylation site"
FT   Modified-site   915..918
FT                   /note= "Amidation site"
FT   Modified-site   947..949
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   981..986
FT                   /note= "N-myristoylation site"
FT   Modified-site   1021..1026
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   1022..1024
FT                   /note= "N-myristoylation site"
FT   Modified-site   1028..1031
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   1031..
FT                   /note= "Casein kinase II phosphorylation site"
XX
PN   WO200140468-A2.
XX
PD   07-JUN-2001.
XX

PF   01-DEC-2000; 2000WO-US32716.
XX
PR   03-DEC-1999; 99US-0168780.
PR   18-FEB-2000; 2000US-0507533.
PR   25-FEB-2000; 2000US-0513904.
PR   10-OCT-2000; 2000US-0685791.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-367809/38.
DR   N-PSDB; AAS05388.
XX
PT   Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10,
PT   CARD-11, useful as targets for therapy, as immunogens, and in screening
PT   and detection assays -
XX
PS   Claim 9; Fig 10A-10C; 145pp; English.
XX
CC   The present sequence represents novel human caspase recruitment
CC   domain, CARD-10. The polynucleotide encoding this sequence was isolated
CC   from a human skin cDNA library. Also described are novel human sequences
CC   for CARD-9 and CARD-11 (AAU01205, AAU01207) and rat CARD-9 (AAU01204).
CC   CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought to
CC   activate nuclear factor (NF)-kappaB and apoptosis. The sequences of the
CC   invention can be used for treating a disorder associated with abnormal
CC   levels of apoptosis by modulating the expression or activity of CARD-9,
CC   CARD-10, or CARD-11. They can be used for the treatment of
CC   hyperproliferative disorders (e.g. cancer), autoimmune disorders (e.g.
CC   systemic lupus erythematosus), neurological disorders (e.g.
CC   Alzheimer's disease), inflammatory disorders (e.g. Crohn's disease),
CC   and viral infection (e.g. HIV). The CARD polypeptide, polynucleotide
CC   and an antibody which selectively binds to CARD can be used in screening
CC   and detection assays (e.g. chromosomal mapping, tissue typing),
CC   predictive medicine (prognostic assays, monitoring clinical trials, and
CC   therapy (treatment and prophylaxis). The CARD polypeptide may be used
CC   to screen for drugs that bind to and/or modulate it. CARD sequences are
CC   potential targets for regulating inflammation, cancer, NF-kappaB
CC   signalling, stress-related response and apoptosis in human disease. A
CC   host cell containing a polynucleotide encoding CARD can be used to
CC   create transgenic animals.
XX
SQ   Sequence 1032 AA;

Query Match           17.23%;  Score 887.5;  DB 22;  Length 1032;
Best Local Similarity 29.2%;   Pred. No. 7.2e-64;
Matches 317;  Conservative 161;  Mismatches 419;  Indels 189;  Gaps 31;

Qy    15 DEETLWEMMESHRHRIVRCICPSRLJTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL 74
          :: ::: ::: :: : : ::::::::::: :: : :::::: : : : :: :::
Db    23 eedalwerleqvrhrlaraInpakTpylrqcrvideqdeeevlstyrfpcrvnrtgrlm 82

Qy    75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGS 134
          ::: ::: ::: :: :: : :::: ::: :: : :: : :
Db    83 dilrcrgkrgyeafiealefyypehtlltggepagrcsmildeegpegltqflmtevrr 142

Qy   135 LQEELNQEKGQKEVLLRRCQQLQEHIGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRL 194
          :: ::: : :: : :: : ::: ::: : : :::: : :::
Db   143 lrearksqlqreqqlqargrvleeeragleqrlrdqqgagercqrlredweagslellrl 202

Qy   195 KDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELEQEQSLRTA 254
          ::: :: : ::: : ::: : : :: : :
Db   203 kdenymiamrlaqlseeknsavlrsrdlqlavdqlklkvsr-----lrra 254

Qy   255 SDQESGDEELNRLKEENE-------KLRS---LTFSLAE------------------- 283
          ::: :: : : ::: ::: : :::
Db   255 rgpppgaeekekekekepdnvdlvselraenqqltasirelqeglqqeasrpgapgse 314

Qy   284 --KDILEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEEKEQTLLQFQKSKMAC 340
          :::: ::: ::: : : :: : ::: :: : : : :: :::
Db   315 rilldilehdwreaqdsrqelcqklhavggelqwaeelrdqylqemedlrlkhrtlqkdc 374
```

Qy 341 QLYREKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELR 400
Db   375 dlykhrmatvlaqleeiekerdqaiqsrdriqlysqsliekdqyrkqvrqleaerdell 434

Qy 401 TOLRQLQAEPPGVLKQEARTR-EPCPREKQRLVRMHAICPRDDSDCSLVSST-------- 451
Db   435 ttltslegtkalllevqlqraqggtclk-----------acasshslcsnlsstwslsefps 484

Qy 452 -------ESQLLSDLSATSSRELVDSFR-----SSSPAPPSQQSLYKRVAEDFGEEP- 496
Db   485 plggpeatgeaavmggpephnseeatdsekeinrlsilpfppsagsilrrqreedpappk 544

Qy 497 WSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVL 556
Db   545 rsfssmsdi--tgsv-------------------tlkpwspglsssssssdsv 575

Qy 557 MRRRPARRILSQVTMLAFQGDALL---------------EQISVIGGNLTGIFIHR 597
Db   576 wplqkpegllargcgldflnrslairvsgrsppgpepqdkgpdglsfygdrwsgavvrr 635

Qy 598 VT--PGSA---ADQMALRPGTQIVMVDYEASEPLFKAVL--EDTTLEEAVGLLRRVDGF 649
Db   636 vlsqpgsarmepreqrveaaglegacleaeaqq--rtllwnqgstlpsl-----mdsk 686

Qy 650 CCLSVKVNTDGYKRLLQDLLEAKVATSG-DSFYIRVNLAMEGRAKGE-LQVHCNEVLHVTD 707
Db   687 acqsfh------------ealeawakgpgaepfyiranltiperadphalcvkageilrivd 736

Qy 708 TMFQCGCGWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSS--G 765
Db   737 saykrrqewfctrvdpltlrd-ldrgtvpnyqraqgll-----evgekcl---pssrhr 786

Qy 766 GPQKLVRIVSMD--------KAKASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTL 818
Db   787 9prsnlkkraldqlrlvrpkpvgpagdspdqillep--------c--aepersrlrpysl 836

Qy 819 VWPHRPARPRPVLLVPRAVGKILSEKLCLQG---FKKCLAEYLSQEEYEAWSQRGDII 874
Db   837 vrpllvsalrpvvllpeclaprlirnlldlpssrldfqvcpaeslsgeelcpssapgapk 896

Qy 875 QEGEVSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKK 934
Db   897 aqpatpglgsri-raiqesvgkk--hcllelgargvrervqnelypivihvevteknvre 953

Qy 935 LKKGLQRLGTSEEQLLEAARQEEGDLDRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQK 994
Db   954 vrgllgrpgwrdsellrqcrgseqvlwglpcswvqvpahewghaelakvvrgrlgeqa 1013

Qy 995 KVVWTE 1000
Db 1014 rlvwve 1019

RESULT 5
AAM93822
ID AAM93822 standard; Protein; 746 AA.
XX
AC AAM93822;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 3879.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.

XX
PR 08-JUL-1999;   99JP-0194486.
PR 11-JAN-2000;   2000JP-018774.
PR 02-MAY-2000;   2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K,   Ishii S,   Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai S, Kojima K, Otsuki T, Koga H;
XX
DR WPI; 2001-52425S/58.
DR N-PSDB; AAK94778.
XX
PT 830 Primers useful for synthesising full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 3879; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 746 AA;

Query Match   11.0%;   Score 566.5;   DB 22;   Length 746;
Best Local Similarity 28.5%;   Pred. No. 1.4e-37;
Matches 223; Conservative 112; Mismatches 300; Indels 147; Gaps 26;

Qy 285 DILEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEKEQTLLQFQKSKMACQLYR 344
Db   33 dilehdwreaqdsrqelcqklhavgqelqwaeelrdqylqemedlrlkhrtlqkdcdlyk 92

Qy 345 EKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLR 404
Db   93 hrmatvlaqleeiekerdqaiqsrdriqlysqsliekdqyrkqvrqleaerdellttlt 152

Qy 405 QLQAEPPGVLKQEARTR-EPCPREKQRLVRMHAICPRDDSDCSLVSST------------ 451
Db 153 slegtkallevqlqraqggtclk-------------acasshslcsnlsstwslsefpsplgg 202

Qy 452 ----ESQLLSDLSATSSRELVDSFR-----SSSPAPPSQQSLYKRVAEDFGEEP-WSFS 500
Db 203 peatgeaavmggpephnseeatdsekeinrlsilpfppsagsilrrqreedpappkrsfs 262

Qy 501 SCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRR 560
Db 263 smsdi--tgsv-------------------tlkpwspglsssssssdsvwplg 293

Qy 561 PARRILSQVTMLAFQGDALL-----------------EQISVIGGNLTGIFIHRVT-- 599
Db 294 kpegllargcgldflnrslairvsgrsppggpepqdkgpdglsfygdrwsgavvrrvlsg 353

Qy 600 PGSA----ADQMALRPGTQIVMVDYEASEPLFKAVL-EDTTLEEAVGLLRRVDGFCCLS 653
Db 354 pgsarmepreqrveaaglegacleaeaqq--rtllwnqgstlpsl-----mdskacqs 404

Qy 654 VKVNTDGYKRLLQDLLEAKVATSG-DSFYIRVNLAMEGRAKGE-LQVHCNEVLHVTDTMFQ 711
Db 405 fh---------ealeawakgpgaepfyiranltiperadphalcvkageilrivdsayk 454

Qy 712 GCGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSS--GGPQK 769
Db 455 rrqewfctrvdpltlrd-ldrgtvpnyqraqgll-----evgekcl----pssrhrgprs 504

```
Qy   770  LVRIVSMD------KAKASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPH  822
            ::   :       ::|:|  |:|  ||:| | ||:||:||-|
Db   505  nlkkraldqlrivrpkpvgapagdspdqlllep-------c--aeperslrpyslvrpl  554

Qy   823  RPARPRPVLLVPRAVGKILSEKLCLLQG----FKKCLAEYLSQEEYEAWSQRGDIIQEGE  878
            |||:|:|:|  : |||||:|||  :||||||||| | ||
Db   555  lvsalrpvvlpeclaprlirnlldlpssrldfqvcpaeslsgeelcpssapgapkaqpa  614

Qy   879  VSGGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNERMAKKLKKG  938
            | | :: ::|||||||||| ::::
Db   615  tpplgsri-raiqesvgkk--hclleigargvrelvqneiypivihvevteknvrevrgl  671

Qy   939  LQRIGTSEEQLLEAARQEGDLDRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVW  998
            | | | ||  | :|| ::||  : : |:|| | ::||
Db   672  lgrpgwrdsellrqcrgseqvlwglpcswvqvpahewghaeelakvvrgrilqeqarlvw  731

Qy   999  TE  1000
            -|
Db   732  ve  733
```

```
RESULT   6
AAU01204
ID   AAU01204 standard; Protein; 536 AA.
XX
AC   AAU01204;
XX
XX
DT   12-SEP-2001 (first entry)
XX
DE   Rat caspase recruitment domain, CARD-9 polypeptide.
XX
KW   Rat; caspase recruitment domain; CARD-9; Bcl-10; NF-kappaB;
KW   apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW   inflammatory disorder; viral infection; stress-related response.
XX
OS   Rattus sp.
XX
FH   Key            Location/Qualifiers
FT   Modified-site  2..5
FT                  /note= "Casein kinase II phosphorylation site"
FT   Domain         7..98
FT                  /note= "CARD domain"
FT   Modified-site  12..15
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  16..18
FT                  /note= "Protein kinase C phosphorylation site"
FT   Modified-site  23..26
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  92..95
FT                  /note= "cAMP- and cGMP-dependent protein kinase
FT                  phosphorylation site"
FT   Modified-site  95..97
FT                  /note= "Protein kinase C phosphorylation site"
FT   Modified-site  95..98
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  138..140
FT                  /note= "Protein kinase C phosphorylation site"
FT   Modified-site  138..141
FT                  /note= "Casein kinase II phosphorylation site"
FT   Domain         140..416
FT                  /note= "Coiled coil domain"
FT   Modified-site  171..174
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  176..183
FT                  /note= "Tyrosine kinase phosphorylation site"
FT   Region         197..213
FT                  /note= "Indole-3-glycerol phosphate synthase
FT                  homology region"
FT   Modified-site  228..231
FT                  /note= "cAMP- and cGMP-dependent protein kinase
FT                  phosphorylation site"
FT   Modified-site  231..233
FT   Modified-site  267..270
FT                  /note= "Protein kinase C phosphorylation site"
FT   Region         285..338
FT                  /note= "Cysteine rich repeat homology region"
FT   Modified-site  303..305
FT                  /note= "Protein kinase C phosphorylation site"
FT   Modified-site  362..364
FT                  /note= "Protein kinase C phosphorylation site"
FT   Modified-site  362..365
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  374..377
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  425..428
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  431..433
FT                  /note= "Protein kinase C phosphorylation site"
FT   Modified-site  451..453
FT                  /note= "Protein kinase C phosphorylation site"
FT   Modified-site  483..486
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  514..516
FT                  /note= "Protein kinase C phosphorylation site"
FT   Modified-site  523..528
FT                  /note= "N-myristoylation site"
FT   Modified-site  524..527
FT                  /note= "N-glycosylation site"
FT   Modified-site  526..529
FT                  /note= "Casein kinase II phosphorylation site"
FT   Modified-site  531..534
FT                  /note= "Casein kinase II phosphorylation site"
XX
PN   WO200140468-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32716.
XX
PR   03-DEC-1999; 99US-0168780.
PR   18-FEB-2000; 2000US-0507533.
PR   25-FEB-2000; 2000US-0513904.
PR   10-OCT-2000; 2000US-0685791.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-367809/38.
DR   N-PSDB; AAS05386.
XX
PT   Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10,
PT   CARD-11, useful as targets for therapy, as immunogens, and in screening
PT   and detection assays -
XX
PS   Claim 9; Fig 1A-1B; 145pp; English.
XX
CC   The present sequence represents novel rat caspase recruitment
CC   domain, CARD-9. The polynucleotide encoding this sequence was
CC   isolated from a rat neuronal cDNA library. Also described are novel
CC   human sequences for CARD-9, CARD-10 and CARD-11 (AAU01205-AAU01207).
CC   CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought
CC   to activate nuclear factor (NF)-kappaB and apoptosis. The sequences
CC   of the invention can be used for treating a disorder associated
CC   with abnormal levels of apoptosis by modulating the expression
CC   or activity of CARD-9, CARD-10, or CARD-11. They can be used for
CC   the treatment of hyperproliferative disorders (e.g. cancer),
CC   autoimmune disorders (e.g. systemic lupus erythematosus), neurological
CC   disorders (e.g. Alzheimer's disease), inflammatory disorders (e.g.
CC   Crohn's disease), and viral infection (e.g. HIV). The CARD polypeptide,
CC   polynucleotide and an antibody which selectively binds to CARD can be
CC   used in screening and detection assays (e.g. chromosomal mapping, tissue
CC   typing), predictive medicine (prognostic assays, monitoring clinical
CC   trials, and therapy (treatment and prophylaxis). The CARD polypeptide may
```

```
CC      be used to screen for drugs that bind to and/or modulate it. CARD
CC      sequences are potential targets for regulating inflammation, cancer,
CC      NF-kappaB signaling, stress-related response and apoptosis in human
CC      disease. A host cell containing a polynucleotide encoding CARD can be
CC      used to create transgenic animals.
XX
SQ      Sequence    536 AA;

Query Match             10.0%;  Score 512.5;  DB 22;  Length 536;
Best Local Similarity   28.8%;  Pred. No. 2.4e-33;
Matches  162;  Conservative   94;  Mismatches  215;  Indels   91;  Gaps  15;

Qy   15 DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL 74
            ::   ||| : ||  |:|||:||||||    ||  :| |||| |::    |  | ||
Db    6 nddecwsalesfrvklisvidpsritpylrqckvlnpddeeqvlsdpnlvirkrkvgvll 65

Qy   75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPD------VDFSNFSGL-------ME 120
         :|: : :|||||||  :  |  |||  | || |   |||    |  | |:       |:
Db   66 dilqrtghkgyvaflesielyypqlyrkvtgkeparvfsmiidasgesgltqllmtevmk 125

Qy  121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
          ::  |  |||  ||:: |    |  :|| |   :  ||  ||               :
Db  126 lqkkvqdltallss-kddfikelrvkdsllrkhqervq------------------rl 164

Qy  181 KREVSAHFHEVLRLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
         |  :  | ||  ||||:||  | |::  |  | || |:||  | ||| | |: ||
Db  165 keecelssaelkrckdenydlamrlahiseekgaalmnrdlqlevdqirhslmkae--d 222

Qy  241 SCELELQEQ-SLRTASDQESGDEELNRLKEENE---KLRSLTFSLAEK-------DI 286
          ||| ::| |:|   :: ||   |||  |:|||    |||  :::: :|       :|
Db  223 dckverkhtlklrhameqrpsqellwdlqgerdllqarvqelevsvqeqklhrnspyiqv 282

Qy  287 LEQSLDEARGSRQELVERIHSLRERAVAAERQREQYWEKEQTLLQFQKSKMACQLYREK 346
         ||  : :   ||||::||:|||:| ||   :::|          ::|||| :|| |:| :
Db  283 leedwrqalqehqeqastifslsrkdlrqaealrrcmeekemfelqclalrkdakmykdr 342

Qy  347 VNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQL 406
          :|   :   |: |||:|    :   | :|| :  ||: |||| || |:  :| |:|:| |
Db  343 ieailqqmeevslerdqamtsreelhaqcaqsfgdkdklrkqvreldekadelqlqlfqt 402

Qy  407 QAEPPGVLKQEARTREPCPREKQRLV---RMHAICPRDDSDCSLVSSTESQL-LSDLSAT 462
          :          :|  | | | |:|     :: | | |  | || |: | || |||: |
Db  403 esr---llaaegrlkq--qqldmlilssdledssprnsqelslpqdleedaqlsdkgyl 456

Qy  463 SSRE-------------LVDSFRSSSSPAPPSQQSLYKRVAEDFG-------EEPWS 498
          ||:             || | |:||: || :|| ||||| |:|        |||
Db  457 adrespeqpfvvinkkhlsqthdtvpssseppekerrrlkesfenyrrkralrkmqnswr 516

Qy  499 FSSCLEIPEGDPGALPGAKAGD 520
          :       ||  |  |:
Db  517 qg------egdhgnttgsdntd 532

RESULT 7
AAU01205
ID   AAU01205 standard; Protein; 536 AA.
XX
AC   AAU01205;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Human caspase recruitment domain, CARD-9 polypeptide.
XX
KW   Human; caspase recruitment domain; CARD-9; Bcl-10; NF-kappaB;
KW   apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW   inflammatory disorder; viral infection; stress-related response.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   2..5
FT                   /note= "Casein kinase II phosphorylation site"
FT   Domain          7..98
FT                   /note= "CARD domain"
FT   Modified-site   23..26
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   92..95
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   95..97
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   95..98
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   138..140
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   138..141
FT                   /note= "Casein kinase II phosphorylation site"
FT   Domain          140..416
FT                   /note= "Coiled coil domain"
FT   Modified-site   176..183
FT                   /note= "Protein kinase C phosphorylation site"
FT   Region          197..213
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Region          228..231
FT                   /note= "Indole-3-glycerol phosphate synthase
FT                   homology region"
FT   Modified-site   228..231
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   231..233
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   267..270
FT                   /note= "Casein kinase II phosphorylation site"
FT   Region          285..338
FT                   /note= "Cysteine rich repeat homology region"
FT   Modified-site   303..305
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   333..336
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   363..366
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   425..428
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   431..433
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   450..452
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   453..458
FT                   /note= "N-myristoylation site"
FT   Modified-site   460..462
FT                   /note= "Protein kinase C phosphorylation site"
FT   Modified-site   481..486
FT                   /note= "N-myristoylation site"
FT   Modified-site   483..486
FT                   /note= "N-myristoylation site"
FT   Modified-site   524..527
FT                   /note= "N-glycosylation site"
FT   Modified-site   526..529
FT                   /note= "Casein kinase II phosphorylation site"
FT   Modified-site   527..532
FT                   /note= "N-myristoylation site"
FT   Modified-site   531..534
FT                   /note= "Casein kinase II phosphorylation site"
XX
PN   WO200140468-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32716.
XX
PR   03-DEC-1999; 99US-0168780.
PR   18-FEB-2000; 2000US-0507533.
PR   25-FEB-2000; 2000US-0513904.
PR   10-OCT-2000; 2000US-0685791.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
```

```
Qy    459  -------LSATSSRELVDSFRS---SSPAPPSQQSLYKRVAEDF 492
           :  :   |:|: ||| :|  || ||::  |: .|| ::  ||::
Db    457  gggspkqpfaalhqevlrnphdaglssgeppeker--rrlkesf 499
```

```
RESULT   8
AAB95617
ID   AAB95617 standard; Protein; 366 AA.
XX
AC   AAB95617;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:18328.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;      99JP-0248036.
PR   27-AUG-1999;      99JP-0300253.
PR   11-JAN-2000;    2000JP-0118776.
PR   02-MAY-2000;    2000JP-0183767.
PR   09-JUN-2000;    2000JP-0241889.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 18328; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesizing 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesizing polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   366 AA;

Query Match          7.78;   Score 396.5;   DB 22;   Length 366;
```

```
XX
PI   Bertin J;
XX
DR   WPI; 2001-367809/38.
DR   N-PSDB; AAS05387.
XX
PT   Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10,
PT   CARD-11, useful as targets for therapy, as immunogens, and in screening
PT   and detection assays -
XX
PS   Claim 9; Fig 5A-5B; 145pp; English.
XX
CC   The present sequence represents novel human caspase recruitment
CC   domain, CARD-9. The polynucleotide encoding this sequence was isolated
CC   from a human megakaryocyte cDNA library. Also described are novel human
CC   sequences for CARD-10 and CARD-11 (AAU01206, AAU01207) and rat CARD-9
CC   (AAU01204). CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is
CC   thought to activate nuclear factor (NF)-kappaB and apoptosis. The
CC   sequences of the invention can be used for treating a disorder
CC   associated with abnormal levels of apoptosis by modulating the
CC   expression or activity of CARD-9, CARD-10, or CARD-11. They can
CC   be used for the treatment of hyperproliferative disorders (e.g.
CC   cancer), autoimmune disorders (e.g. systemic lupus erythematosus),
CC   neurological disorders (e.g. Alzheimer's disease), inflammatory disorders
CC   (e.g. Crohn's disease), and viral infection (e.g. HIV). The CARD
CC   polypeptide, polynucleotide and an antibody which selectively binds to
CC   CARD can be used in screening and detection assays (e.g. chromosomal
CC   mapping, tissue typing), predictive medicine (prognostic assays,
CC   monitoring clinical trials, and therapy (treatment and prophylaxis). The
CC   CARD polypeptide may be used to screen for drugs that bind to and/or
CC   modulate it. CARD sequences are potential targets for regulating
CC   inflammation, cancer, NF-kappaB signalling, stress-related response and
CC   apoptosis in human disease. A host cell containing a polynucleotide
CC   encoding CARD can be used to create transgenic animals.
XX
SQ   Sequence   536 AA;

Query Match          9.6%;        Score 493;   DB 22;   Length 536;
Best Local Similarity   29.5%;   Pred. No. 9.9e-32;
Matches  155;  Conservative  92;  Mismatches  200;  Indels  78;  Gaps  14;
```

```
Qy     15  DEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRLTNSAMRAGHLL 74
               : :  || : |  ||||:||||||  .||:||||   :||| :|| | |
Db      6  nddecwnvlegfrvtltsvidpsritpylrgckvlnpddeeqvlsdpnlvirkrkvgvll 65

Qy     75  DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPD-------VDFSNFSGL-------ME 120
           |:| :|| |||:|||:|:|||| :|:|| || |:|        | : ||:|       ||
Db     66  dilqrtghkgyvafleslelyypglykkvtgkeparvfsmiidasgesgltgllmtevmk 125

Qy    121  TSKLTECLAGATGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
           : :|:|:    |||:| :|: |||| |: ||: :  ||:|:|:              :|
Db    126  lqkkvqdltaliss-kddfikelrvkdsllrkhqervq-------------rl 164

Qy    181  KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
             |  | | |  :|   || ::  | : ||:||||| :||||  |||  |:| |
Db    165  keeceagsrelkrckeenydlamrlahqseekgaalmrndlqleidqlkhslmkae--d 222

Qy    241  SCELELEQ-SLRTASDQESGDEELNRLKEE----NEKLRSLTFSLAEK--------DI 286
           |  |:|   ||  :||| |::|:|||:|:|     ||  :|: |  |:|        :|
Db    223  dckverkhtlklrhameqrpsqellwelqqekallqarvqeleasvqegkldrsspyiqv 282

Qy    287  LEQSLDEARGSRQELVERIHSLRERAVAAEREQRQYWEKEQTLLQFQKSKMACQLYREK 346
           | :.:| |  :||: :|:  :|  |. :|:||:|   |: |  |:|  ||| ||  |  :
Db    283  leedwrqalrdhqeqantifsirkdlrqgearrlrcmeekemfelqclalrkdskmykdr 342

Qy    347  VNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQL 406
           :| |   : :  :: | |   : : :|  |:  |  :||||||:|   :|  ||  |:||
Db    343  ieaillqmeevaierdqaiatreelhaqharglqekdalrkqvrelgekadelqlqvfqc 402

Qy    407  QAEPPGVLKQEARTREPCPREKQRLVRMHAI----CPRDDSDCSLVSSTESQLLSD------ 458
           :|  |  |||:| | :|   || | |:|        |||| |  || | ||| |:|
Db    403  eaq--llavegrlrr---qqietlvissdledgsprrsqelslpqdledtqlsdkgcla 456
```

```
Best Local Similarity 30.7%;  Pred. No. 5e-24;
Matches 116;  Conservative 65;  Mismatches 146;  Indels 51;  Gaps 8;

Qy   15 DETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRLTNSAMRAGHLL 74
            ||   :|:|   || :||||:|||||| |||   ||||    ||
Db    6 ndecwnvlegfrvtltsvidpsritpylrgckvlnpddeeqvlsdpnlvirkrkvgvll 65

Qy   75 DLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPD------VDFSNFSGL------ME 120
         |:| |||||||:||||||  | ||    |   ||         |:   || |||         |:
Db   66 dilrgthkgyvafleslelyypglykkvtgkeparvfsmidasgesgltqllmtevmk 125

Qy  121 TSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRM 180
           | :|  |||: ||||||||||||  ||||:||| ||:||||||||||||||||:    |:
Db  126 lqkkvqdltaliss-kddfikelrvkdsllrkhqervq------------------rl 164

Qy  181 KREVSAHFHEVLRLKDEMLSLSLHYSNALQEKELAASRCRSIQEELYLLKQELQRANMVS 240
              :||      :|| :||       :||   : |
Db  165 keeceagsrelkrckeenydlanrlahqseekgaalmrndlqleidglkhslmkae--d 222

Qy  241 SCELELQEQ-SLRTASDQESGDEELNRLKEE----NEKLRSLTFSLAEK------DI 286
              :||    |   ||        ||          ||      :|       |  ||
Db  223 dckverkhtlkirhamegrpsqellwelqgekallqarvqeleasvqegkldrssspyiqv 282

Qy  287 LEQSLDEARGSRQELVERIHSLRERAVAAERQRQYWEEKEQTLLQFQKSKMACQLYREK 346
            |: :|||      ::||  ||||        :|||: ||    ::  :::
Db  283 leedwrqalrdhqeqantifslrkdlrqgeaarrlrcmeekemfelqclalrkdskmykdr 342

Qy  347 VNALQAQVCELQKERDQA 364
            ||   ||||  ||||
Db  343 ieaillqmeevalerdqs 360
```

RESULT 9
AAM38934
ID   AAM38934 standard; Protein; 350 AA.
XX
AC   AAM38934;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 2079.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR   WPI; 2001-442253/47.

DR   N-PSDB; AAI58090.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 3; SEQ ID NO 2079; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 350 AA;

```
Query Match          6.0%;  Score 309;  DB 22;  Length 350;
Best Local Similarity 32.2%;  Pred. No. 8e-17;
Matches 112;  Conservative 50;  Mismatches 148;  Indels 38;  Gaps 12;

Qy  668 LEAKVATSG-DSFYIRVNLAMEGRAKGE-LQVHCNEVLHVTDTMFQGCGCWHAHRVNSYT 725
          |||   || ||| |||  | ||| |||| : ||   ::  | |
Db   13 leawakgpgaepfyiranltlperadphalcvkageilrlvdsaykrrqewfctrvdplt 72

Qy  726 MKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSS--GGPQKLVRIVSMD------ 777
          ::| :|||| |||| || ||  ::  ||    ||       ::
Db   73 lrd-ldrgtvpnyqraqqll-----evqekcl---pssrhrgprsnlkkraldqlrlvr 122

Qy  778 -KAKASPLRLSFDRGQLDPSRMEGSSTCFWAESCLTLVPVTLVWPHRPARPRPVLLVPRA 836
           |:::|   :| ||||:|||| |   |   :||| |: ||||||  |   |||||  :|
Db  123 pkpvgapagdspdqillep------c--aepersirpysylvrplvsalrpvvllpec 172

Qy  837 VGKILSEKLCLLQG---FKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTRHAVE 892
          :  | :|  |||       ||   ||  |  :|  ||  :||     || || ||
Db  173 laprlirnlldlpssrldfqvcpaeslsgeelcpssapgakqpatpgigsri-raige 231

Qy  893 SLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGTSEEQLLEA 952
          |:    : |  ||     ::|| ||||||:||:||:   :|:: |     : ||
Db  232 svgkk--hcllelgargvrelvqneiyplvihvevteknvrevrgllgrpgwrdsellrq 289

Qy  953 ARQEEGDLDRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWTE 1000
           |    |   :|   ||   ||  :  |   || :|  | :||| |||
Db  290 crgseqvlwglpcswvqvpahewghaeelakvvrgrilqegqarlvwve 337
```

RESULT 10
ABB64619
ID   ABB64619 standard; Protein; 1445 AA.
XX
AC   ABB64619;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 20649.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO20017042-A2.
XX

```
PD    27-SEP-2001.
XX
PF    23-MAR-2001; 2001WO-US09231.
XX
PR    23-MAR-2000; 2000US-191637P.
PR    11-JUL-2000; 2000US-0614150.
XX
PA    (PEKE ) PE CORP NY.
XX
PI    Venter JC, Adams M, Li PWD, Myers EW;
XX
DR    WPI; 2001-656860/75.
DR    N-PSDB; ABL08722.
XX
PT    New isolated nucleic acid detection reagent for detecting 1000 or more
PT    genes from Drosophila and for elucidating cell signalling and cell-cell
PT    interactions -
XX
PS    Disclosure; SEQ ID NO 20649; 21pp + Sequence Listing; English.
XX
CC    The invention relates to an isolated nucleic acid detection reagent
CC    capable of detecting 1000 or more genes from Drosophila. The invention is
CC    useful in developmental biology and in elucidating cell signalling and
CC    cell-cell interactions in higher eukaryotes for the development of
CC    insecticides, therapeutics and pharmaceutical drugs. The invention
CC    discloses genomic DNA sequences (ABL01840-ABL30511), expressed DNA
CC    sequences (ABL6176-ABL16175) and the encoded proteins
CC    (ABB57737-ABB72072).
CC    The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence    1445 AA;

Query Match        5.5%;  Score 282;  DB 22;  Length 1445;
Best Local Similarity  22.7%;  Pred. No. 1.3e-13;
Matches  155; Conservative 111; Mismatches 276; Indels 142; Gaps 30;

Qy    373  REISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLV  432
             : ||| :  :  |: |:    : ||: || |:    ||: : ||: : || |:
Db    194  keissk--areqlnangysl--qegdiitrihntncgdtmslkeakkiidgc-kerlnlv  248

Qy    433  RMHAICPRDD-SDCSLVSSTESQLLSDLSAT----------SSRELVDSFRSSSPAPPSQ  481
             : :  |  |||:|  || : | :|  |: :          || |: |: :|
Db    249  vlrditnqtavsqlnlnnsashqasgniyathqpqvsgcssnnnledpylpgg-asyss  307

Qy    482  QSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSLQ  541
             : ||:  |  |  : :  | :  |  |  :|  ||  |:||||||: |:  |:
Db    308  qnlyvqpptrtsngpningngln--deksnltprgrsrgp------imdgvslqqld---r  357

Qy    542  PVSPGR-----LDV--------SESGVLM-------RRRPARRILSQVTMLAFQGDALL  580
             || :     |:|        :||: |        : :| : : ::||:  : :
Db    358  pvtptrgsaaideprprPpPPrgssgaaqedfyssrrqlyeerqsaeprfisfqkegsv  417

Qy    581  EQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTLEEAV  640
             :| | | ||   | || |||: || | |: || : || |:|:: | :|| |:|: ||
Db    418  -giirltggneagifvtavgpgspasiqglmpgdkilkvn-------dmdmnqvtreeav  468

Qy    641  GLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVATS--GDSFYIRVNLAMEGRAKGELQVH  698
             | | | :| :|: :  |: ||: |  : |: | |    |:: :|: ::|| | |:|:|
Db    469  lfl------lslqdridlivqyckeeydevvtnqrgdsfhikthfhcdnpskgemafk  520

Qy    699  CNEVLHVTDTMFQG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCT  757
             ::|: ::| :| |  ||| |  | |  : : : | || || :|  || |: |: |: |:
Db    521  agdvfrvidtlhngvvgswqvlkigrqhqe-mqrgvipnksraeel------ataqfn  571

Qy    758  VTRKPSSGGPQKLVRIVSMDKAKASPLRLSFDRGQLDPSR------MEGSSTCFWAESCL  811
             |:| |   |    ||:| :  |  |||| :  |:|| ||        |||::| | :||
Db    572  atkk------------emnanesrgnfrrrrsthrrskslsrenwddvvfsdsis  615

Qy    812  TLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRG  871
             | | | |   |:||||:||  :  |||| ||:||: |||: : || ||:|||:|| | |
Db    4    LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL  63
```

```
Db    616  kfpayervvlrhgfvrpvvlf-gpvsdliarerla--kdfpdkfstplqddksa----  667
             |:| ||| || |: ||: |  |||: : |: ||    |::|: ::|| :::
Qy    872  DIIQEGEVSGGRCWVTRHA-VESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEK  930
             : ||  ::|: : | :|:  || | |:|  || |:|| :| |||: |: ||:| || |
Db    668  ------atsgkcrivrlsnirdvmdrgkhalldltpnavdrlnyaqfypvviflktdsk  720
XX
Qy    931  -MAKKLKKGLQRLG-TSEEQLLEAARQEEGLDRAPCLYSSLAPDGW----------SDL  978
             :| :| :| |||| || :  ||   |: ::| :  | : ||| ||          |
Db    721  hvikqlrhglpkaahkssklleqcqk----lerv-----------wshifstqialsde  765
XX
Qy    979  DGLLSCVRQAIADEQKKVVWTEQS  1002
             : :| :  :  :  :  :  ::| :|:
Db    766  eswyrklrdsidlqsgavwmses  789
```

```
RESULT 11
AAR47173
ID    AAR47173 standard; Protein; 2101 AA.
XX
AC    AAR47173;
XX
DT    15-JUL-1994 (first entry)
XX
DE    Sequence of the inner nuclear matrix (INM) protein MT2.
XX
KW    Malignant cell type marker; interior nuclear matrix; MT2; NuMA.
XX
OS    Homo sapiens.
XX
PN    WO9440573-A.
XX
PD    06-JAN-1994.
XX
PF    21-JUN-1993;   93WO-US06160.
XX
PR    22-JUN-1992;   92US-0901701.
XX
PA    (MATR-) MATRITECH INC.
XX
PI    Lidgard GP, Toukatly G;
XX
DR    WPI; 1994-026210/03.
DR    N-PSDB; AAQ54841.
XX
PT    Novel malignant cell type markers of the interior nuclear matrix
PT    - used for detecting abnormal cell types e.g. malignant breast,
PT    prostate, lung, etc., for determining deg. of cell death in
PT    tissue, etc.
XX
PS    Claim 39; Page 56-66; 93pp; English.
XX
CC    MT1 and MT2 proteins were isolated from malignant cells essentially
CC    following the method of Penman and Fey, described in US Pat. Nos. 4,
CC    882,268 and 4,885,236. DNA sequences encoding these proteins were
CC    cloned by probing expression libraries using monoclonal antibodies
CC    raised against the isolated INM proteins. The genetic sequence
CC    encoding MT2 has been disclosed by Yang et al. 1992, J. Cell Biol.
CC    116:1395-1408, and is referred to by them as 'NuMA'. The nucleotide
CC    sequence encoding MT1 has not beed described previously. Both the
CC    MT1 and MT2 proteins are distributed throughout the nucleus (with
CC    the exception of the nucleolus) in non-mitotic cells, and localise
CC    to the spindle during mitosis. A binding protein having a binding
CC    affinity of greater than about 10(9) M(-1) for AAR47173 is claimed
CC    for use in the manufacture of a medicament.
XX
SQ    Sequence    2101 AA;

Query Match        5.1%;  Score 262;  DB 15;  Length 2101;
Best Local Similarity  19.9%;  Pred. No. 1.1e-11;
Matches  208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

Qy    4    LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL  63
```

```
AAG65799
ID   AAG65799 standard; protein; 2101 AA.
XX
AC   AAG65799;
XX
DT   30-JAN-2002 (first entry)
XX
DE   Nuclear mitotic apparatus protein (NuMA) sequence.
XX
KW   NuMA; Nuclear Mitotic Apparatus protein; malignant; cell growth; human;
KW   tumor progression; structural protein.
XX
OS   Homo sapiens.
XX
PN   US6287790-B1.
XX
PD   11-SEP-2001.
XX
PF   30-NOV-1999;   99US-0452294.
XX
PR   30-NOV-1998;   98US-110420P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Lelievre S, Bissell M;
XX
DR   WPI; 2001-610603/70.
XX
PT   Distinguishing growth arrested malignant cells from proliferating
PT   non-malignant cells, comprises staining specifically the nuclear mitotic apparatus
PT   protein in intact mammalian cells and determining the 3-dimensional
PT   pattern of the labeled protein -
XX
PS   Disclosure; Fig 15; 38pp; English.
XX
CC   The invention relates to methods of distinguishing malignant or growth
CC   arrested malignant cells from proliferating non-malignant cells by
CC   staining specifically Nuclear Mitotic Apparatus (NuMA) protein in a
CC   sample of intact mammalian cells and imaging the cells to determine the
CC   three dimensional pattern of labeled NuMA protein within nucleic of the
CC   cells. The method is useful for distinguishing malignant or growth
CC   arrested malignant cells from proliferating non-malignant cells.
CC   Localization of structural proteins such as NuMA may be used to identify
CC   tumors cells and different stages in the tumor progression and
CC   differentiation processes. NuMA proteins may be used to transport
CC   functional proteins or drugs to specific locations inside the cells.
CC   The present sequence represents the human NuMA protein (GenBank Accession
CC   Number Z11584).
XX
SQ   Sequence  2101 AA;
```

```
Query Match             5.1%;  Score 262;  DB 22;  Length 2101;
Best Local Similarity  19.9%;  Pred. No. 1.1e-11;
Matches  208;  Conservative  163;  Mismatches  376;  Indels  296;  Gaps  39;
```

```
Qy    4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL 63

Db  263 llnekqaasplepkeleelrdkn----------esltmrlhetlkgcqdlkteksqmdrk 312

Qy   64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS 122

Db  313 inqlse---------engdisfklrefashlqqlqdalnelteehskatqewlekqa 360

Qy  123 KLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182

Db  361 qlekelsaa--lqdkkclee-kneilqgklsqleehls----------qlqdnppqpekq 406

Qy  183 EVSAHFHEVLRLKDEMLSLS-----------LHYSNALQEKELAASRCRSLQEELYL-- 228

Db  407 evlgdvlqletlkqeaatlaanntqlqarvemletergqqeakllaerghfeeekqqlss 466

Qy  229 -----------LKQELQRANMVSSCELEQEQSLRT------ASDQESGDEELNRLK 268

Db  467 litdlqssisnlsqakeeleqasahgarltaqvaslteltinatigqq-dqelagik 525

Qy  269 EE------------------NEKLRSLTFSLABKDILEQSLDEARGSRQELVERIH 306

Db  526 qqakeqaqlaqtlqqgeqasqglrhqveqlsssikqk---eqqlkev-aekqeatrqdh 581

Qy  307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS 366

Db  582 a-qqlataae-ereaslrerdaalkqle-----alekekaaklei----lqgqlqvane 629

Qy  367 ARDSAQREISQSLVEKDSLRRQVFEL-----------TDQVCELRTQLRQLQAEPPG 412

Db  630 ardsaqtsvtqagrekaelsrkveelqacvetargeqheaqaqvaelelqlrseq----- 684

Qy  413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472

Db  685 --qkateKervaqekdqi------------------qeqlqalkeslivtkgsleekr 723

Qy  473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532

Db  724 raadaleeqqrciselkae-------trslveghkrerkeleeeragrkglearllqige 776

Qy  533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRIISQVTMLAFQGDALLEQISVIGGNLTG 592

Db  777 ahqaete-------------vlrrelaeamaaqht----aeseceql----------- 806

Qy  593 IFIHRVTPGSAADOMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652

Db  807 ---------vkevaawrdgye---dsqoeeaqyganfqeq----------------1m 836

Qy  653 SVKVNTDGYKRLLQDLEARKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712

Db  837 tlkeecekargelqeakekva--gieshselqisrgqnklael--han----------- 880

Qy  713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLV- 771

Db  881 ---laralqgvqekevraqkladdlstlqekmaatskevarletlvrk--ageqqetas 934

Qy  772 RIVSMDKAKASPLRLSFDRGQLDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR 828

Db  935 relvkepara-----------gdrpewleeqgrqfc------------------stqa 965

Qy  829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTR 888

Db  966 alqanereaeqmgne---lerlraalmesgqgqqeerqgqerevarltqerg------r 1015

Qy  889 HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI-VIHVSVNEKMAK------KLK--- 936

Db 1016 aqdalalekaaraelemriqnalneqrvefatiqealahalteKegkdqelakIrgleaa 1075

Qy  937 --KGLQRLGTSEEQLLEAARQEE 957

Db 1076 qlkeleelrqtvkqlkeqlakke 1098
```

RESULT 12

```
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US08656.
XX
PR   18-APR-2000; 2000US-0552929.
PR   26-JAN-2001; 2001US-0770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy -
XX
PS   Claim 20; Page 547; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising
CC   the nucleic acids encoding the polypeptides and cells genetically
CC   engineered to express them are also useful for producing the proteins.
CC   The proteins are useful in genetic vaccination, testing and
CC   therapy, and can be used as nutritional supplements. They may be used to
CC   increase stem cell proliferation; to regulate haematopoiesis; and in
CC   bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC   immune suppression and/or stimulation; as anti-inflammatory agents; and
CC   in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC   sequences of novel human secreted proteins of the invention.
XX
SQ   Sequence   2207 AA;

Query Match              5.1%;   Score 262;  DB 22;  Length 2207;
Best Local Similarity   19.9%;   Pred. No. 1.1e-11;
Matches  208;  Conservative 163;  Mismatches 376;  Indels 296;  Gaps 39;

Qy      4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL 63
                    :  :       :  :   :  :   ::: ::   :   :::   : :  ::
Db    269 llnekqaasplepkeleelrdkn---------esltmrlhetlkgcqdlkteksqmdrk 318

Qy     64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS 122
                 :: ::  ::: :  :: ::   ::   :  :  :::  : :: :::
Db    319 inqlse--------engdlsfklrefashlqqlqdalnelteehskatqewlekqa 366

Qy    123 KLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182
                   : :  ::    :      ::  :: :   :::         :::   ::
Db    367 qlekelsaa--lqdkkclee-kneilgqklsqleehls------------qlqdnppqekg 412

Qy    183 EVSAHFHEVLRLKDEMLSLS--------------------LHYSNALQEKELAASRCRSLQEELYL-- 228
             ::                                    :   :: ::::   : :   ::  ::
Db    413 evlgdviqletlkqeeatlaanntqlqarvemletergqqeakllaerghfeeekqqlss 472

Qy    229 ------------LKQELQRANMVSSCELELQEQSLRT----ASDQESGDEELNRLK 268
                        ::  :  :::     ::     :::     :: : ::::  ::
Db    473 litdlqssinlsqakeleqasqahgarltaqvasltselttlnatiqqg-dqelaglk 531

Qy    269 EE----------------------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
             ::                            ::  ::  : :  ::   :::   :: :::
Db    532 qqakekqaqlaqtlqgqeqasqglrhqveqlssslkqk--eqqlkev-aekqeatrqdh 587

Qy    307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS 366
                :: ::   ::   :  :  :       :     :  :  :  :::  :  :::::
Db    588 a-qqlataae-ereaslrerdaalkqle-----alekekaaklei----lqqlqvane 635

Qy    367 ARDSAQREISQSLVEKDSLRRQVFEL------------------TDQVCELRTQLRQLQAEPPG 412
```

```
Db    467 litdlqssinlsqakeleqasqahgarltaqvasltselttlnatiqqg-dqelaglk 525
                 :::: :::   :::      ::      :    ::    : ::  :::   ::
Qy    269 EE----------------------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
             ::                            ::  :   :  ::   :::   :: :::
Db    526 qqakekqaqlaqtlqgqeqasqglrhqveqlssslkqk--eqqlkev-aekqeatrqdh 581

Qy    307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS 366
                :: ::   ::   :  :  :       :     :  :  :  :::  :  :::::
Db    582 a-qqlataae-ereaslrerdaalkqle-----alekekaaklei----lqqlqvane 629

Qy    367 ARDSAQREISQSLVEKDSLRRQVFEL------------------TDQVCELRTQLRQLQAEPPG 412
             :::: ::  ::  :  :  ::::  :                 :  :
Db    630 ardsaqtsvtgaqrekaelsrkveelqacvetarqeqheaqavaelelqlrseq------ 684

Qy    413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472
             : : :  :  :    :                                 :  :  :
Db    685 --qkatekervaqekdql---------------qelgalkeslkvtkgsleeekr 723

Qy    473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
                 :   :       :      ::      :                 :   :: ::
Db    724 raadaleeggrciselkae------------trslveqhkrerkeleeeragrkglearlliqlqe 776

Qy    533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
               : :           :  :      : :   :  : :::    : : :
Db    777 ahqaete---------vlrrelaeamaaqht----aeseceql--------- 806

Qy    593 IFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652
                : :         :     :       ::         : :    :  :  :   :
Db    807 ----------vkevaawrdgye---dsqqeeaygamfgeq--------------lm 836

Qy    653 SVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712
                 :   :  : :     ::  :  :     :     ::   :   :   ::    ::
Db    837 tlkeecekargelqeakekva-gleshselqisrqqnklael--han--------- 880

Qy    713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIODMTQQCTVTRKPSSGGPQKLV- 771
             : :       :           :  :    :      :   :     ::: :::
Db    881 ---laralqqvqekevraqkladdistlqekmaatskevarletlvrk--ageqqetas 934

Qy    772 RIVSMDKAKASPLRLSFDRGQLDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR 828
                 : : :        ::  :   :  :            :   :  ::     :
Db    935 relvkepara-------gdrqpewleeqgrqfc---------------stqa 965

Qy    829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQDEGEVSSGRCWVTR 888
             : :  :           :  :: :::   :   ::        :        :   :
Db    966 alqamereaemgne----lerlraalmesqgqqeergqqerevarltqerg-----r 1015

Qy    889 HAVESLMBKNTHALLDVOLDSVCTLHRMDIFPI--VIHVSVNEKMAK----KLK------ 936
                 ::          ::   :                 :::             ::
Db   1016 aqadlaieka araelemriqnalneqrvefatlqeaiahaltekegkdqelakirgleaa 1075

Qy    937 --KGLQRLGTSEEQLLEAARQEE 957
                : :  :      :: ::
Db   1076 qikeleelrqtvkqlkeqlakke 1098

RESULT 13
AAU32041
ID   AAU32041 standard; Protein; 2207 AA.
XX
AC   AAU32041;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #2532.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO20017944 9-A2.
```

PA   (UJYO) UNIV JOHNS HOPKINS.
PA   (MERI) MERCK FROSST CANADA INC.
XX
PI   Thornberry N,   Rosen A,   Casciola-Rosen L,   Andrade FA,   Nicholson D;
PI   Roy S;
XX
DR   WPI; 2000-013162/01.
XX
PT   Autoantigenic fragments useful for diagnosis, treatment of autoimmune
PT   diseases and cancer -
XX
PS   Claim 4; Fig 9; 88pp; English.
XX
CC   The present invention describes autoantigenic fragments (I) produced by
CC   the action of a lymphocyte granule enzyme (II) on isolated cells
CC   containing an autoantigen. (I) is useful for prophylactic and therapeutic
CC   treatment of an autoimmune disease (AI) such as type I diabetes mellitus,
CC   thyroiditis, myasthenia gravis, primary biliary cirrhosis, systemic lupus
CC   erythematosus, rheumatoid arthritis, polymyositis, dermatomyositis,
CC   Sjogren's syndrome, scleroderma and graft-vs-host disease. (I) produced
CC   by the action of (II) on cells isolated from target tissue is
CC   administered to a patient for tolerising to the presence of (I). As a
CC   therapeutic treatment, isolated (I) associated with AI condition is
CC   contacted with serum of the patient containing autoantibodies (AAb)
CC   against (I), under condition to allow the in vivo binding of AAb to (I)
CC   and a portion of the AAb is removed from the serum of the patient. (I)
CC   derived from malignant cell is administered to a patient for stimulating
CC   an immune response against malignant cells to treat cancer. (I) is also
CC   useful in assays for assessing the presence or absence of an AI condition
CC   in a patient. Such assays involve detecting AAb in a sample by using (I)
CC   or detecting (I) in a sample by using antibodies that specifically binds
CC   to a cryptic epitope of (I). The present sequence represents the human
CC   NuMA protein, which is a specifically claimed autoantigen from the
CC   present invention.
XX
SQ   Sequence   2101 AA;

Query Match   5.0%;   Score 259;   DB 21;   Length 2101;
Best Local Similarity   19.9%;   Pred. No. 1.9e-11;
Matches 208;   Conservative 162;   Mismatches 377;   Indels 296;   Gaps 39;

Qy   4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRITPYLRQAKVLCQLDEEVLHSPRL 63
                                                              ::::
Db   263 llnekqaasplepkeelrdkn---------------esltmrlhetlkqcqdikteksqmdrk 312

Qy   64 TNSAMRAGHLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS 122
            || :| :| ::| :::
Db   313 inqlse---------------engdlsfklrefashlqglqdalnelteehskatqewlekqa 360

Qy   123 KLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182
           || :: | :|| | :::| | ||:|||
Db   361 qlekelsaa---lgdkkclee-kneilqgkisqleehls---------qlqdnppgekg 406

Qy   183 EVSAHFHEVLRLKDEMLSLS------------------LHYSNALQBEKELAASRCRSLQEELYL-- 228
            || :||:|
Db   407 evlgdvlqletlkqeaatlaanntqlgarvemletergqeakllaerghfeeekqqlss 466

Qy   229 ---LKQELQRANMVSSCELELQEQSLRT------ASDQESGDEELNRLK 268
              ||:|: |||: : | :||:|: :| |: ::||
Db   467 litdlqssisnlsqakeeleqasqahgarltaqvasltselttlnatiqqq-dqelaglk 525

Qy   269 EE------------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
            ::   :| :: |::| :| :|
Db   526 qqakekqaqlaqtlqqgeqasqqlrhqveqlsssslkqk--eqqlkev-aekqeatrqdh 581

Qy   307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS 366
                                       |||: :::|| ||: ||
Db   582 a-qqlatsae-ereaslrerdaalkqle-----alekekaaklei---lqgqlqvane 629

Qy   367 ARDSAQREISQSLVEKDSLRRQVFEL----------TDQVCELRTQLRQLQAEPPG 412
            |||| |||: :: ||:| ||||| :| | | | : ||
Db   630 ardsaqtsvtgaqrekaelsrkveelgacvetarqgeheagaqvaelelqlrseq----- 684

RESULT 14
AAY49936
ID   AAY49936 standard; protein; 2101 AA.
XX
AC   AAY49936;
XX
DT   01-FEB-2000 (first entry)
XX
DE   Human NuMA protein #1.
XX
KW   Human; NuMA; intranuclear protein; mitosis; segregation; DNA-PK-cs;
KW   DNA-activated protein kinase; catalytic subunit; PARP; detection;
KW   nuclear NAD+ ADP-ribosyltransferase; autoantigen; immune response;
KW   autoimmune disease; cancer; type I diabetes mellitus; thyroiditis;
KW   myasthenia gravis; primary biliary cirrhosis; rheumatoid arthritis;
KW   systemic lupus erythematosus; polymyositis; dermatomyositis;
KW   Sjogren's syndrome; scleroderma; graft-vs-host disease.
XX
OS   Homo sapiens.
XX
PN   WO9953757-A1.
XX
PD   28-OCT-1999.
XX
PF   22-APR-1999; 99WO-US08774.
XX
PR   22-APR-1998; 98US-0082643.
XX
PA   (MERI) MERCK & CO INC.

Db   636 ardsaqtsvtgaqrekaelsrkveelgacvetarqgeheagaqvaelelqlrseq----- 690
              ||||| ::: || : | |:| || ||| |||
Qy   413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472
           |:   :: ||
Db   691 --qkatekrvaqekdql-------------------qeqlqalkeslkvtkgsleeekr 729

Qy   473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
            :|| :::| :|
Db   730 raadaleeqqrciselkae------trslveqhkrerkeleeragrkglearllqlge 782

Qy   533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
              ||:   :::   :|:: ||
Db   783 ahqete---------------vlrrelaeamaaght------aesec eql--------- 812

Qy   593 IFIHRVTPGSAADQMALRPGTQIMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652
                    :::| : : | ::   ::
Db   813 -----------vkevaawrdgye---dsqeeaygamfqeq--------------lm 842

Qy   653 SVKVNTDGYKRLLQDLEARKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712
                || |:: | : : :|| || |:
Db   843 tlkeecekarqelqeakekva-gieshselqisrqgnklael--han----------- 886

Qy   713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQLIALIQDMTQQCTVTRKPSSGGPQKLV- 771
            :   :|| :::|| |:: :: : || :: | | |:
Db   887 ----laralqgvqekevraqkladdlstlqekmaatskevarletlvrk--ageqqetas 940

Qy   772 RIVSMDKAKASPLRLSFDRGQLDPSRME--GSSTCFWAESCLTLVPYTLVWPHRPARPR 828
             |: :: | |
Db   941 relvkepara--------------gdrqpewleeqggrqfc--------------stqa 971

Qy   829 PVLLVPRAVGKIILSEKLICLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTR 888
             :   :|:   ::| | || |
Db   972 alqamereaeqmgne---leriraalmesggqqeergqqerevaritqerg------r 1021

Qy   889 HAVESLMEKNTHALLLDVQLDSVCTLHRMDIFPI-VIHVSVNEKMAK----KLK------ 936
            |: || ::: ||:
Db   1022 aqadialekaaraelemrlqnalneqrvefatlqealahaltekegkdqelakirgleaa 1081

Qy   937 --KGLQRLGTSEEQLLEAARQEE 957
               ::   ::   :|
Db   1082 qikeleelrqtvkqlkeqlakke 1104

PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 14061; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511), and the encoded proteins
CC   (ABB5737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1916 AA;

Query Match   5.0%;   Score 258;   DB 22;   Length 1916;
Best Local Similarity   22.1%;   Pred. No. 2e-11;
Matches   133; Conservative   97; Mismatches 209; Indels 164; Gaps   27;

Qy   474 SSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDY--------- 525
           ::|| || | : :| || | ||| | :||
Db   1402 nsprpparnslfp-----------------lpmqpqapstrpgsraplshqsikdqsftds 1445

Qy   526 ----ELLDTADLPQLESSLQPVSPGRLDVSESGVLMRRPARRI------------LSQV 569
               :|:| ::| ::| | :::|||
Db   1446 lengsdisssqdmpssaattttt----asatstvydeepkpslppppasvpaetlryv 1500

Qy   570 TMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKA 629
           || |: |||:| | |:| || || : :: |: ::| ||| |:   :::|
Db   1501 tlhmdkskn1--giklfggnkvgiyvhdvavgspsdhagirkgdqi--leyngvd----- 1551

Qy   630 VLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQD---LEAKVATSGDSFYIRVNL 685
            |||:|: | | :| ::: |: || | | :|||| |
Db   1552 -lsgvtaeqaaneiskl------------------tdtvtmlvqnklhtlkqikdepgdsfyirvgf 1599

Qy   686 AMEGRA-KGELQVHCNEVLHVTDTMFQGC-GCWHAHRVNSYTMKDTAAH---GTIPNYS 739
             : :: || | | | :| || |: :| :::
Db   1600 drtgelneddirfvkdevlyydntvfngtfglwrawkl------damghrkecgiipsqm 1653

Qy   740 RAQQQL----IALIQDMTQQCTVTRKPSSGGPQKLVRIVSMDK---AKASPLRLSF--- 788
           | ||| | || || | || : || :|||:| | | || :| |
Db   1654 kveeelrsgevvdcdtgtarrgstsarrsfirrknqrsssrdsteiasfsntqlsffpd 1713

Qy   789 ------DRGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILS 842
                     :   :| : |:|| |
Db   1714 lglinddgga1syqrvellds---------------------pir-rpvliigp-----ls 1747

Qy   843 E---------KLCLLQGF--KKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGGRCW 885
           | |:| | :: | : | :: | || | ::|
Db   1748 eclmvrltidfsnlfklcevtamdcsqeameegikenifvdyrrrgnkfe------ct 1799

Qy   886 VTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQRLGTS 945
           | : | : | |||||: :|| | | || |: ||:   :||| ||| |||
Db   1800 tveaisnackndrrhcildvsisaverlqrlqiyplvl--lirfksakqi-rdirdfgtd 1856

Qy   946 E---------EQLLEAARQEEGDLDRAPCLYSSLAPDGWSDLDGLLSCVRQAIADEQKKVVWT 999
           | :|:|| | || : :| | ||:|: ::: || |: |||| |
Db   1857 kisakaakemyeramkletdykq---yisavipgvs-ikhmctqikdavdkeqkllwv 1911

Qy   1000 EQS 1002
           |
Db   1912 pvs 1914

RESULT 15
ABB62423
ID   ABB62423 standard; Protein; 1916 AA.
XX
AC   ABB62423;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 14061.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-655860/75.
DR   N-PSDB;-ABL06526.
XX

Qy   413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472
           |:| :|| :| ::|| ::|
Db   685 --qkatekervaqekdql-----------------qeqlqalkeslkvtkgsleeekr 723

Qy   473 SSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
           :: || | : :: || | | ::
Db   724 raada1eeqqrciselkae------trslveghkrerkeleeeragrkglearllqlge 776

Qy   533 LPQLESSLQPVSPGRLDVSESGVLMRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
                         ::|| : || :: ||
Db   777 ahqaete----------------vlrrelaeamaaqht----aeseceql--------- 806

Qy   593 IFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652
                                                 |:
Db   807 -------vkevaawrdgye---dsqqeeaqygamfgeq------------------lm 836

Qy   653 SVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712
                                                       | | |
Db   837 tikeeecekargelqeakekva--gieshselqisrqqnklael--han---------- 880

Qy   713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGGPQKLV- 771
             : : |::: || || : |: ||: || | | :: | |
Db   881 ---laralgqvqekevraqkladd1stlqekmaatskevarletlvrk--sgeqqetas 934

Qy   772 RIVSMDKAKASPLRLSFDRGQLDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR 828
           |:| | | :||
Db   935 relvkepara---------gdrqpewleeqgrqfc------------------stga 965

Qy   829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGGRCWVTR 888
            : :| ::| | ::| :| | | : | :
Db   966 alqamereaegmgne----lerlraalmesqgqqeergqqerevarltqerg-------r 1015

Qy   889 HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI--VIHVSVNEKMAK----KLK------ 936
                  :| |: |: |:| :: | || |:|
Db   1016 aqadialekaaraelemrlqnalneqrvefatlqealahaltekegkdqelaklrglesa 1075

Qy   937 --KGLQRLGTSEEQLLEAARQEE 957
             : | :| :|||
Db   1076 qikeleelrqtvkqlkeglakke 1098

Search completed: June 13, 2002, 09:23:52
Job time: 235 sec

```
             GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    June 13, 2002, 09:21:47 ; Search time 17.94 Seconds
                                     (without alignments)
                                     1366.962 Million cell updates/sec

Title:           US-09-767-215-2
Perfect score:   5149
Sequence:        1 MGELCRRDSALTALDEETLW.........VRQAIADEQKKVWTEQSPR 1004

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        231628 seqs, 2425594 residues

Total number of hits satisfying chosen parameters:    231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Issued_Patents_AA:*
         1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
         2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
         3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
         4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
         5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
         6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	262	5.1	2101	1	US-08-466-390-4	Sequence 4, Appli
2	262	5.1	2101	1	US-08-470-950-4	Sequence 4, Appli
3	262	5.1	2101	1	US-08-467-781-4	Sequence 4, Appli
4	262	5.1	2101	2	US-08-195-487-4	Sequence 4, Appli
5	262	5.1	2101	4	US-08-483-924-4	Sequence 4, Appli
6	262	5.1	2101	1	US-09-452-294-1	Sequence 1, Appli
7	262	4.6	3248	5	PCT-US93-06160-4	Sequence 4, Appli
8	238	4.6	3248	5	US-08-353-702-1	Sequence 1, Appli
9	238	4.6	3248	5	PCT-US95-16216-1	Sequence 1, Appli
10	237	4.6	2482	2	US-08-328-254-6	Sequence 6, Appli
11	216.5	4.2	1388	2	US-08-685-576-1	Sequence 1, Appli
12	210.5	4.1	976	4	US-09-100-324B-4	Sequence 4, Appli
13	205	4.0	816	2	US-08-533-306A-6	Sequence 6, Appli
14	205	4.0	816	2	US-08-742-923A-6	Sequence 6, Appli
15	205	4.0	885	2	US-08-533-306A-4	Sequence 4, Appli
16	205	4.0	885	2	US-08-742-923A-4	Sequence 4, Appli
17	204.5	4.0	1354	3	US-08-685-871-2	Sequence 2, Appli
18	200.5	3.9	1388	2	US-08-685-576-4	Sequence 4, Appli
19	197	3.8	576	2	US-08-533-306A-2	Sequence 2, Appli
20	197	3.8	576	2	US-08-742-923A-2	Sequence 2, Appli
21	197	3.8	1886	4	US-08-938-105-3	Sequence 3, Appli
22	195	3.8	1090	5	US-09-085-199B-5	Sequence 5, Appli
23	194	3.8	1388	4	US-09-572-191-2	Sequence 2, Appli
24	192.5	3.7	756	9	US-09-085-199B-9	Sequence 9, Appli
25	190	3.7	914	4	US-09-085-199B-4	Sequence 4, Appli
26	189	3.7	1898	1	US-08-056-200-94	Sequence 94, Appl
27	189	3.7	1898	2	US-08-800-644-94	Sequence 94, Appl
28	189	3.7	1939	4	US-09-310-187A-1	Sequence 1, Appli
29	186.5	3.6	955	1	US-08-006-676B-1	Sequence 1, Appli
30	186.5	3.6	955	1	US-08-282-845-2	Sequence 2, Appli
31	186.5	3.6	955	5	PCT-US94-00324-1	Sequence 2, Appli
32	186	3.6	683	6	5210183-3	Patent No. 5210183
33	182.5	3.5	955	6	US-08-428-414A-3	Sequence 3, Appli
34	181	3.5	1618	1	US-07-853-913-4	Sequence 4, Appli
35	176.5	3.4	900	2	US-08-630-822A-62	Sequence 62, Appl
36	176.5	3.4	900	2	US-09-005-069-62	Sequence 62, Appl
37	176	3.4	1162	2	US-08-728-323A-2	Sequence 2, Appli
38	175	3.4	835	4	US-08-968-751-4	Sequence 4, Appli
39	175	3.4	1068	4	US-09-085-199B-11	Sequence 11, App
40	174.5	3.4	712	2	US-08-468-576B-17	Sequence 17, App
41	174.5	3.4	712	3	US-08-468-579B-17	Sequence 17, App
42	174.5	3.4	712	3	US-08-468-577B-17	Sequence 17, App
43	173.5	3.4	1375	4	US-09-722-139-2	Sequence 2, Appli
44	166.5	3.2	1312	2	US-08-687-080-51	Sequence 51, App
45	165.5	3.2	829	1	US-07-670-611-2	Sequence 2, Appli

```
                              ALIGNMENTS

RESULT  1
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
;   APPLICANT: TOUKATLY, GARY
;   APPLICANT: LIDGARD, GRAHAM P
;   TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;   TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;     STREET: 125 HIGH STREET
;     CITY: BOSTON
;     STATE: MA
;     COUNTRY: USA
;     ZIP: 02110
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC Compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/466,390
;     FILING DATE: 06-JUN-1995
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: PITCHER ESQ, EDMUND R
;     REGISTRATION NUMBER: 27,829
;     REFERENCE/DOCKET NUMBER: MTP-013
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (617) 248-7000
;     TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2101 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-466-390-4

Query Match         5.1%;  Score 262;  DB 1;  Length 2101;
Best Local Similarity  19.9%;  Pred. No. 5.7e-13;
Matches  208;  Conservative 163;  Mismatches 376;  Indels 296;  Gaps 39;

Qy     4  LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL 63
             |.:|::|:.|:|:||  :| | ||    |:| ||::|       || .| ||
Db   263  LLNEKQAASPLEPKELEELRDKN------ESLTMRLHETLKQCODLKTEKSQMDRK 312
```

```
GENERAL INFORMATION:
  APPLICANT: TOUKATLY, GARY
  APPLICANT: LIDGARD, GRAHAM P
  TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
  TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
  NUMBER OF SEQUENCES: 6
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: TESTA, HURWITZ & THIBEAULT
    STREET: 125 HIGH STREET
    CITY: BOSTON
    STATE: MA
    COUNTRY: USA
    ZIP: 02110
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/470,950
    FILING DATE: 06-JUN-1995
    CLASSIFICATION: 536
  ATTORNEY/AGENT INFORMATION:
    NAME: PITCHER ESQ, EDMUND R
    REGISTRATION NUMBER: 27,829
    REFERENCE/DOCKET NUMBER: MTP-013
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (617) 248-7000
    TELEFAX: (617) 248-7100
  INFORMATION FOR SEQ ID NO: 4:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 2101 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-470-950-4
```

```
Query Match            5.1%;   Score 262;  DB 1;   Length 2101;
Best Local Similarity  19.9%;  Pred. No. 5.7e-13;
Matches  208;  Conservative 163;  Mismatches 376;  Indels 296;  Gaps

Qy    4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRL  63
Db  263 LLNEKQAASPLEPKELEELRDKN-------------ESLTMRLHETLKQCQDLKTEKSQMDRK 312

Qy   64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS 122
Db  313 INQLSE------------ENGDLSFKLREFASHLQQLQDALNELTEEHSKATQEWLEKQA  360

Qy  123 KLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182
Db  361 QLEKELSAA--LQDKKCLEE-KNEILQGKLSQLEEHLS------------QLQDNPPQEKG  406

Qy  183 EVSAHFHEVLRLKDEMLSLS-----------LHYSNALQEKELAASRCRSLQEELYL-- 228
Db  407 EVLGDVLQLETLKQEAATLAANNTQLQARVEMLETERGQQEAKLLAERGHFEEEKQQLSS  466

Qy  229 ----------LKQELQRANMVSSCELELQEQSLRT-----ASDQESGDEELNRLK     268
Db  467 LITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELTTLNATIQQQ-DQELAGLK  525

Qy  269 EE------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH          306
Db  526 QQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQK---EQQLKEV-AEKQEATRQDH  581

Qy  307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS 366
Db  582 A-QQLATAAE-EREASLRERDAALKQLE-----LQQQLQVANE                  629

Qy  367 ARDSAQREISQSLVEKDSLRRQVFEL------TDQVCELRTQLRQLQAEPPG         412
Db  630 ARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAVAELELQLRSEQ------ 684
```

```
Qy  413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472
Db  685 --QKATEKERVAQEKDQL-----------QEQLQALKESLKVTKGSLEEEKR         723

Qy  473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
Db  724 RAADALEEQQRCISELKAE-----TRSLVEQHKRERKELEEERAGRKGLEARLLQLGE   776

Qy  533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
Db  777 AHQAETE----------------------VLRRELAEAMAAQHT-----AESECEQL-   806

Qy  593 IFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652
Db  807 -----------VKEVAAWRDGYE---DSQQEEAQYGAMFQEQ---------------LM  836

Qy  653 SVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712
Db  837 TLKEECEKARQELQEAKEKVA--GIESHSELQISRQQNKLAEL-HAN----------   880

Qy  713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVRKPSSGGPQKLV-  771
Db  881 ---LARALQQQVQEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRK---AGEQQETAS 934

Qy  772 RIVSMDRKAKASPLRLSFDRGQLDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR 828
Db  935 RELVKEPARA--------GDRQPEWLEEQQGRQFC-------------STQA         965

Qy  829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTR 888
Db  966 ALQAMEREAEQMGNE---LERLRAALMESQGQQQEERGQQEREVARLTQERG----R  1015

Qy  889 HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI-VIHVSVNEKMAK----KLK-      936
Db 1016 AQADLALEKAAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKIRGLEAA 1075

Qy  937 -KGLQRLGTSEEQLLEAARQEE                                       957
Db 1076 QIKELEELRQTVKQLKEQLAKKE                                      1098
```

```
RESULT 2
US-08-470-950-4
; Sequence 4, Application US/08470950
;.Patent No. 5698439
```

```
;       TELEPHONE:  (617) 248-7000
;       TELEFAX:    (617) 248-7100
;  INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;       LENGTH: 2101 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;       MOLECULE TYPE: protein
US-08-467-781-4
```

```
Query Match             Best Local Similarity   19.9%;
                        Score 262;  DB 1;   Length 2101;
Matches 208;  Conservative 163;  Mismatches 376;  Indels 296;  Gaps 39;
                        Pred. No. 5.7e-13;  5.1%;
```

```
Qy    4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL   63
          | |  : |:::  |:   | |:|
Db  263 LLNEKQAASPLEPKELEELRDKN------------ESLTMRLHETLKQCQDLKTEKSQMDRK  312

Qy   64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS  122
          :  | | | | |  |
Db  313 INQLSE------------------ENGDLSFKLREFASHLQQLQDALNELTEHSKATQEWLEKQA  360

Qy  123 KLTECLAGAIGSLQELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR  182
          |:: :|:     |:       ::     |||
Db  361 QLEKELSAA---LQDKKCLEE-KNEILQGKLSQLEEHLS------------QLQDNPPQEKG  406

Qy  183 EVSAHFHEVLRLKDEMLSLS----------LHYSNALQEKELAASRCRSLQEELYL-  228
          :|  |  |  |       |
Db  407 EVLGDVLQLETLKQEAATLAANNTQLQARVEMLETERGQQEAKLLAERGHFEEEKQQLSS  466

Qy  229 ------------LKQELQRANMVSSCELELQEQSLRT------ASDQESGDEELNRLK  268
                      |  | |      : :
Db  467 LITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELTTLNATIQQQ-DQELAGLK  525

Qy  269 EE----------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH  306
          ::
Db  526 QQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQK--EQQLKEV-AEKQEATRQDH  581

Qy  307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS  366
          :: :|||:|| :||  : : |      |::
Db  582 A-QQLATAAE-EREASLRERDAALKQLE------ALEKEKAAKLEI---LQQQLQVANE  629

Qy  367 ARDSAQREISQSLVEKDSLRRQVFEL------------TDQVCELRTQLRQLQAEPPG  412
          ||||:| ::||:  |::||:|| :|
Db  630 ARDSAQTSVTQAQREKAELSRKVEELQACVETARQEHEAQAVAELELQLRSEQ------  684
```

```
Qy  413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR  472
          :|| |   : | |:                          ||:|::|:||
Db  685 --QKATEKERVAQEKDQL---------------------------QEQLQALKESLKVTKGSLEEEKR  723

Qy  473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD  532
          ::|| ||                |     || |     | ||  |  ||
Db  724 RAADALEEQQRCISELKAE-----TRSLVEQHKRERKELEEERAGRKGLEARLLQLGE  776

Qy  533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG  592
          |   |                                         ||:
Db  777 AHQAETE------------------------VLRRELAEAMAAQHT----AESECEQL------  806

Qy  593 IFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL  652
          : :|                                      :|
Db  807 --------------VKEVAAWRDGYE----DSQQEEAQYGAMFQEQ--------------LM  836

Qy  653 SVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG  712
          ::|  |  |: ||: |  |  |: |   :   :  : |  |
Db  837 TLKEECEKARQELQEAKEKVA--GIESHSELQISRQQNKLAEL--HAN------------  880

Qy  713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLV-  771
          |                 |:::  |    :|      :  :|: : |
Db  881 ----LARALQQVQEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRK--AGEQQETAS  934

Qy  772 RIVSMDKAKASPLRLSFDRGQLDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR  828
          |  |   :    |
Db  935 RELVKEPARA--------GDRQPEWLEEQQGRQFC--------------------STQA  965

Qy  829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSSGGRCWVTR  888
          :|  |     |  : : :|     |  |  |:
Db  966 AlQAMEREAEQMGNE----LERLRAALMESGGQQQEREVARLTQERG------R 1015

Qy  889 HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI--VIHVSVNEKMAK----KLK-----  936
          : :||   |   |  ::| |       |     ||  :|
Db 1016 AQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKLRGLEAA 1075

Qy  937 --KGLQRLGTSEEQLLEAARQEE  957
          :|  |||       |  ::|
Db 1076 QIKELEELRQTVKQLKEQLAKKE 1098
```

```
Db  935 RELVKEPARA--------GDRQPEWLEEQQGRQFC-----------------STQA 965
         |:.: ::::|
Qy  829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTR 888
         |::|||| || ::  |: |:::|  |  |::
Db  966 ALQAMEREAEQMGNE----LERLRAALMESQGQQQEERGQQEREVARLTQERG------R 1015
         :||::|  ||  |   ||:|
Qy  889 HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI--VIHVSVNEKMAK----KLK----- 936
         :|:|                |||  :: ||
Db 1016 AQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKLRGLEAA 1075

Qy  937 --KGLQRLGTSEEQLLEAARQEE 957
         |||:|:|||:  :||
Db 1076 QIKELEELRQTVKQLKEQLAKKE 1098

RESULT  4
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
;  APPLICANT: TOUKATLY, GARY
;  APPLICANT: LIDGARD, GRAHAM P
;  TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;  TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;  NUMBER OF SEQUENCES:  6
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  TESTA HURWITZ & THIBEAULT
;   STREET:  53 STATE STREET
;   CITY:  BOSTON
;   STATE:  MA
;   COUNTRY:  USA
;   ZIP:  02109
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/195,487
;   FILING DATE:
;   CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  US/07/901,701
;   FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
;   NAME:  PITCHER ESQ, EDMUND R
;   REGISTRATION NUMBER:  27,829
;   REFERENCE/DOCKET NUMBER:  MTP-013
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  617/248-7000
;   TELEFAX:  617/248-7100
; INFORMATION FOR SEQ ID NO:  4:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH:  2101 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;  MOLECULE TYPE:  protein
US-08-195-487-4

Query Match           5.1%;  Score 262;  Pred. No. 5.7e-13;
Best Local Similarity  19.9%;  Pred. No. 5.7e-13;
Matches  208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

Qy    4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRL 63
         |  |:|:|   ||  :| |  ||
Db  263 LLNEKQAASPLEPKELEELRDKN--------ESLIMRLHETLKQCQDLKTEKSQMDRK 312
Qy   64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVTLVTGLQPDVDFSNFSGLMETS 122
         | :::||   :    |::|:|
Db  313 INQLSE----------ENGDLSFKLREFASHLQQLQDALNELTEEHSKATQEWLEKQA 360
Qy  123 KLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182
```

RESULT 5
US-08-483-924-4
; Sequence 4, Application US/08483924
; Patent No. 5882876
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX

```
Db  361 QLEKELSAA--LQDKKCLEE-KNEILQGKLSQLEEHLS---------------QLQDNPPQEKG 406
         :|:|:|    |:.: ::::|      |::|:|
Qy  183 EVSAHFHEVLRLKDEMLSLS--------------LHYSNALQEKELAASRCRSLQEELYL-- 228
         ||                    || : |:  ||:   :|: |
Db  407 EVLGDVLQLETLKQEAATLAANNTLQARVEMLETERGQQEFAKLLAERGHFEEEKQQLSS 466
         || : :||  ::|| |
Qy  229 -------------------LKQELQRANMVSSCELELQEQSLRT-----ASDQESGDEELNRLK 268
         :|
Db  467 LITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELTTLNATIQQQ-DQELAGLK 525
         |
Qy  269 EE------------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
         ::                  |:|:|   ||  |:   :|
Db  526 QQAKEKQAQLAQTLQQQEQASQLRHQVEQLSSSLKQK--EQQLKEV-AEKQEATRQDH 581
         |
Qy  307 SLRERAVAABRQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAVCELQKERDQAYS 366
         : :: |||| |||  :|| :|:      |: :||
Db  582 A-QQLATAAE-EREASLRERDAALKQLE---ALEKEKAAKLEI---LQQQLQVANE 629
Qy  367 ARDSAQREISQSLVEKDSLRRQVFEL----------------TDQVCELRTQLRQLQAEPPG 412
         ||||| ||:|                         |:|  |::|
Db  630 ARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQQAVAELELQLRSEQ------ 684
Qy  413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472
         |:|  |:|   ::|
Db  685 --QRATEKERVAQEKDQL--------------QEQLQALKESLKVTKGSLEEEKR 723
Qy  473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
         :|     |     |||  |:  |  |:
Db  724 RAADALEEQQRCISELKAE------TRSLVEQHKRERKELEEERAGRKGLEARLLQLGE 776
Qy  533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
         :||  :   ||
Db  777 AHQAETE-------------------VLRRELAEAMAAQHT----AESECEQL------ 806
Qy  593 IFIHRVTPGSAADQMALRPGTQIVMDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652
         |  |:    |   :|:      :  :
Db  807 ----------VKEVAAWRDGYE----DSQQEEAQYGAMFQEQ----------LM 836
Qy  653 SVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712
         :|| ::|  :: : |    ||::
Db  837 TLKEECEKARQELQEAKEKVA--GIESHSELQISRQQNKLAEL--HAN-------- 880
Qy  713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQOLIALIQDMTQQCTVTRKPSSGGPQKLV- 771
         | :|     |:   || || |||     :|
Db  881 ----LARALQQVQEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRK--AGEQQETAS 934
Qy  772 RIVSMDKAKASPLRLSFDRGQLDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR 828
         |:  : :::|                    : |||:
Db  935 RELVKEPARA---------GDRQPEWLEEQQGRQFC-----------------STQA 965
Qy  829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTR 888
         |::|||| || ::  |: |:::|  |  |::
Db  966 ALQAMEREAEQMGNE---LERLRAALMESQGQQQEERGQQEREVARLTQERG------R 1015
         :||::|  ||  |   ||:|
Qy  889 HAVESLMEKNTHALLDSVCTLHRMDIFPI--VIHVSVNEKMAK----KLK----- 936
         :|:|                |||  :: ||
Db 1016 AQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKLRGLEAA 1075
Qy  937 --KGLQRLGTSEEQLLEAARQEE 957
         |||:|:|||:  :||
Db 1076 QIKELEELRQTVKQLKEQLAKKE 1098
```

```
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-924-4
```

```
Query Match           5.1%; Score 262; DB 2; Length 2101;
Best Local Similarity 19.9%; Pred. No. 5.7e-13;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

Qy    4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL 63
Db  263 LLNEKQAASPLEPKELEELRDKN-----------ESLIMRLHETLKQCQDLKTEKSQMDRK 312

Qy   64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS 122
Db  313 INQLSE----------ENGDLSFKLREFASHLQQLQDALNELTEEHSKATQEWLEKQA 360

Qy  123 KLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182
Db  361 QLEKELSAA--LQDKKCLEE-KNEILQGKLSQLEEHLS---------QLQDNPPQEKG 406

Qy  183 EVSAHFHEVLRLKDEMLSLS----------LHYSNALQEKELAASRCRSLQEELYL-- 228
Db  407 EVLGDVLQLETLKQEAATLAANNTQLQARVEMLETERGQQEAKLLAERGHFEEEKQQLSS 466

Qy  229 --------LKQELQRANMVSSCELELQEQSLRT-------ASDQESGDEELNRLK 268
Db  467 LITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELTTLNATIQQQ-DQELAGLK 525

Qy  269 EE------------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
Db  526 QQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQK--EQQLKEV-AEKQEATRQDH 581

Qy  307 SLRERAVAAERQEYWEEKEQTLLQFOKSKMACQLYREKVNALQAQVCELQKERDQAYS 366
Db  582 A-QQLATAAE-EREASLRERDAALKQLE-----ALEKEKAAKLEI---LQQQLQVANE 629

Qy  367 ARDSAQREISQSLVEKDSLRRQYFEL-----------TDQVCELRTQLRQLQAEPPG 412
Db  630 ARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAVAELELQLRSEQ------ 684

Qy  413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472
Db  685 ---QKATEKERVAQEKDQL------------QEQLQALKESLKVTKGSLEEEKR 723

Qy  473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
Db  724 RAADALEEQQRCISELKAE------TRSLVEQHKRERKELEEERAGRGKGLEARLLQLGE 776

Qy  533 LPQLESSLQPVSPGRLDVSESGVLMRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
Db  777 AHQAETE--------------------VLRRELAEAMAAQHT-----AESECEQL-- 806

Qy  593 IFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652
Db  807 -----------VKEVAAWRDGYE----DSQQEEAQYGAMFQEQ-------------LM 836

Qy  653 SVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712
Db  837 TLKEECEKARQELQEAKEKVA--GIESHSELQISRQQNKLAEL--HAN---------- 880

Qy  713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLV- 771
Db  881 ---LARALQQVQEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRK-AGEQQETAS 934

Qy  772 RIVSMDKAKASPLRLSFDRGQLDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR 828
Db  935 RELVKEPARA--------GDRQPEWLEEQQGRQFC---------------STQA 965

Qy  829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTR 888
Db  966 ALQAMEREAEQMGNE---LERLRAALMESGQQQEERGQQEREVARLTQERG------R 1015

Qy  889 HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI-VIHVSVNEKMAK----KLK--- 936
Db 1016 AQADILALEKAAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKLRGLEAA 1075

Qy  937 --KGLQRLGTSEEQLLEAARQEE 957
Db 1076 QIKELEELRQTVKQLKEQLAKKE 1098
```

```
RESULT 6
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
; APPLICANT: Lelievre, Sophie
; APPLICANT: Bissell, Mina
; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
; TITLE OF INVENTION: THERAPH AND DETECTION OF PROLIFERATIVE AND
; TITLE OF INVENTION: DIFFERENTIATION DISORDERS
; FILE REFERENCE: IB-1454- Sequence Submittal
; Patent No. 6287790
; CURRENT APPLICATION NUMBER: US/09/452,294
; CURRENT FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: 60/110,420
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-452-294-1
```

```
Query Match           5.1%; Score 262; DB 4; Length 2101;
Best Local Similarity 19.9%; Pred. No. 5.7e-13;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

Qy    4 LCRRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEVLHSPRL 63
Db  263 LLNEKQAASPLEPKELEELRDKN-----------ESLIMRLHETLKQCQDLKTEKSQMDRK 312

Qy   64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS 122
Db  313 INQLSE----------ENGDLSFKLREFASHLQQIQDALNELTEEHSKATQEWLEKQA 360
```

```
;;   NUMBER OF SEQUENCES:  6
;;   CORRESPONDENCE ADDRESS:
;;      ADDRESSEE: TESTA HURWITZ & THIBEAULT
;;      STREET:  53 STATE STREET
;;      CITY:  BOSTON
;;      STATE:  MA
;;      COUNTRY:  USA
;;      ZIP:  02109
;;   COMPUTER READABLE FORM:
;;      MEDIUM TYPE:  Floppy disk
;;      COMPUTER:  IBM PC compatible
;;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;;   CURRENT APPLICATION DATA:
;;      APPLICATION NUMBER:  PCT/US93/06160
;;      FILING DATE:  19930621
;;      CLASSIFICATION:
;;   ATTORNEY/AGENT INFORMATION:
;;      NAME:  PITCHER ESQ, EDMUND R
;;      REGISTRATION NUMBER:  27,829
;;      REFERENCE/DOCKET NUMBER:  MTP-013
;;   TELECOMMUNICATION INFORMATION:
;;      TELEPHONE:  617/248-7000
;;      TELEFAX:  617/248-7100
;;   INFORMATION FOR SEQ ID NO: 4:
;;      SEQUENCE CHARACTERISTICS:
;;         LENGTH:  2101 amino acids
;;         TYPE:  amino acid
;;         TOPOLOGY:  linear
;;      MOLECULE TYPE:  protein
;;  PCT-US93-06160-4
```

```
Query Match             5.1%;  Score 262;  DB 5;  Length 2101;
Best Local Similarity   19.9%;  Pred. No. 5.7e-13;
Matches  208;  Conservative 163;  Mismatches 376;  Indels 296;  Gaps 39;

Qy     4 LCRRDSALTALDEETLWEMEESHRHRIVRCICPSRLTPYLRQAKVLCQLDEEEVLHSPRL 63
           ||||| :: ||:  ::|| :| |  ||::  :|
Db   263 LLNEKQAASPLEPKELEELRDKN--------ESLTMRLHETLKQCQDLKTEKSQMDRK 312

Qy    64 TNSAMRAGHLLDLLKTRGKNGAIAF-LESLKFHNPDVYTLVTGLQPDVDFSNFSGLMETS 122
            ||: ::  | :|  :::    || |
Db   313 INQLSE--------ENGDLSFKLREFASHLQQLQDALNELTEEHSKATQEWLEKQA 360

Qy   123 KLTECLAGATGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182
           :: : :|| |: :|:  |  ||::: |||:|| |:||| | ||:| ||||
Db   361 QLEKELSAA----LQDKKCLEE-KNEILQGKLSQLEEHLS-----------QLQDNPPQEKG 406

Qy   183 EVSAHFHEVLRLKDEMLSLS--------LHYSNALQEKELAASRCRSLQEELYL- 228
           || |: |: :|  :: |:    :| |:|  :|
Db   407 EVLGDVLQLETLKQEAATLAANNTQLQARVEMLETERGQQEAKLLAERGHFEEEKQQLSS 466

Qy   229 ---------LKQELQRANMVSSCELELQEQSLRT-----ASDQESGDEELNRLK 268
                    |::| ||  ||: |: :: |:|     |:|  : ||:|
Db   467 LITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELTTLNATIQQQ-DQELAGLK 525

Qy   269 EE---------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
           ::              :|:|| || :|: :|::  : | : :||:|
Db   526 QQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQK--EQQLKEV-AEKQEATRQDH 581

Qy   307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS 366
           :: :|| || |||  || || |:| ||  : | | |::: |  :|:|  ::
Db   582 A-QQLATAAE-EREASLRERDAALKQLE----------LQQQLQVANE 629

Qy   367 ARDSAQREISQSLVEKDSLRRQVFEL--------TDQVCELRTQLRQLQAEPPG 412
           |||||| :|:|: ::| ||:|:||  || || || |:| ||:|
Db   630 ARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQ------ 684

Qy   413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472
           | ::| :||  |::| ||:  : |: | | :: |
Db   685 ---QKATEKERVAQEKDQL----------QEQLQALKESLKVTKGSLEEEKR 723
```

```
Qy   123 KLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKR 182
           :: : :|| |: :|:  |  ||::: |||:|| |:||| | ||:| ||||
Db   361 QLEKELSAA--LQDKKCLEE-KNEILQGKLSQLEEHLS-----------QLQDNPPQEKG 406

Qy   183 EVSAHFHEVLRLKDEMLSLS--------LHYSNALQEKELAASRCRSLQEELYL-- 228
           || |: |: :|  :: |:    :| |:|  :|
Db   407 EVLGDVLQLETLKQEAATLAANNTQLQARVEMLETERGQQEAKLLAERGHFEEEKQQLSS 466

Qy   229 ---------LKQELQRANMVSSCELELQEQSLRT-----ASDQESGDEELNRLK 268
                    |::| ||  ||: |: :: |:|     |:|  : ||:|
Db   467 LITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELTTLNATIQQQ-DQELAGLK 525

Qy   269 EE---------------NEKLRSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
           ::              :|:|| || :|: :|::  : | : :||:|
Db   526 QQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQK--EQQLKEV-AEKQEATRQDH 581

Qy   307 SLRERAVAAERQREQYWEEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYS 366
           :: :|| || |||  || || |:| ||  : | | |::: |  :|:|  ::
Db   582 A-QQLATAAE-EREASLRERDAALKQLE----------LQQQLQVANE 629

Qy   367 ARDSAQREISQSLVEKDSLRRQVFEL--------TDQVCELRTQLRQLQAEPPG 412
           |||||| :|:|: ::| ||:|:||  || || || |:| ||:|
Db   630 ARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQ----- 684

Qy   413 VLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSRELVDSFR 472
           | ::| :||  |::| ||:  : |: | | :: |
Db   685 ---QKATEKERVAQEKDQL----------QEQLQALKESLKVTKGSLEEEKR 723

Qy   473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
           :: :   :    |::   :|    :| |: | |   | ::
Db   724 RAADALEEQQRCISELKAE------TRSLVEQHKREKELEEERAGRKGLEARLLQLGE 776

Qy   533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
           : : :: :    :     ::|:|  || |  ||
Db   777 AHQAETE----------VLRRELAEAMAAQHT----AESECEQL----- 806

Qy   593 IFIHRVTPGSAADQMALRPGTQIVMVDVEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652
                     : :|   ||  : | |    ::  ||  |:
Db   807 --------VKEVAAWRDGYE--DSQQEEAQYGAMFQEQ------LM 836

Qy   653 SVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712
              |:    ||  |||||:||   |   |   ||   |
Db   837 TLKEECEKARQELQEAKEKVA-GIESHSELQISRQQNKLAEL-HAN------ 880

Qy   713 CGGWHAHRVNSYTMKDTAAHGTIPNYSRAQQLIALIQDMTQQCTVFRKPSSSGPQKLV- 771
              |     :::     | :   | :| :  | :| : |: | :
Db   881 ----LARALQQVQEKEVRAQKLADDLSTIQEKMAATSKEVARLETLVRK--AGEQQETAS 934

Qy   772 RIVSMDKAKASPLRLSFDRGQLQDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR 828
             | |    |    || :|  |       |
Db   935 RELVKEPARA----------GDRQPEWLEEQQGRQFC--------STQA 965

Qy   829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYFAWSQRGDIIQEGEVSGGRCWVTR 888
              : :   :   |::   ||    :: | ||  :   ||
Db   966 ALQAMEREAEQMGNE---LERLRAALMESQGQQQEEERGQQQEEREVARLTQERG-----R 1015

Qy   889 HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI-VIHVSVNEKMAK---KLK------ 936
             | : ::|:    |:  :|:    |:     :|   : | :::|:
Db  1016 AQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKLRGLEAA 1075

Qy   937 --KGLQRLGTSEEQLLEAARQEE 957
               | |  ::|::|| |:|
Db  1076 QIKELEELRQTVKQLKEQLAKKE 1098
```

```
RESULT 7
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT:
;  TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;  TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
```

```
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: HUMAN
US-08-353-700-1

Query Match          4.6%;     Score 238;  DB 1;   Length 3248;
Best Local Similarity  19.4%;   Pred. No. 1.3e-10;
Matches  214;  Conservative 192;  Mismatches 393;  Indels 306;  Gaps 48;

Qy     2 GELCRRDSALTALD------EETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDE 54
           ||  | ||     :  :||                       :|
Db  2061 GELDTMSKKTALDQLSERMKET-QELESHQSECLHCI---------------QVAE 2102

Qy    55 EEVLHSPRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVTLVTGLQPDVDFSN 114
Db  2103 AEVKEKTELLQTL--SSDVSELLKD-----------KTH--LQEKLQSLEKD- 2139

Qy   115 FSGLMETSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLG------ 161
Db  2140 ---SQALSLTKC------ELENQTAQLNKEKELLVKESESLQARLSESDYEKLNVSKAL 2189

Qy   162 ------------------LAETRAEGLHQLEADHSRMKREVSA--------HFHEVLR----L 194
Db  2190 EAALVEKGEFALRLSSTQEE-VHQLRRGIEKLRVRIEADEKKQLHIAEKLKERERENDSL 2248

Qy   195 KDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCEL-ELQEQSLRT 253
Db  2249 KDKVENLERELQMSEENQELVILDAENSKAEVETLKTQIE--EMARSLKIFELDLVTLRS 2306

Db   254 ASDQ--ESGDEELNRLKEENEKLRSLTFSLAEKDILE-QSLDEARGSRQELVERIHSLRE 310

Db  2307 EKENLTKQIQEKOQGLSELDKLLSSFKSLLEEKEQAEIQIKEESKTAVEMQNQLKELNE 2366

Qy   311 RAVAAERQREQYWEEKEQTL----------------------LQFQKSKMACQLYREK---- 346
Db  2367 -AVAALCGDQEIMKATEQSLDPPIEEEHQLRNSIEKLRARLEADEKKQLCVLQQLKESEH 2425

Qy   347 -VNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQ 405
Db  2426 HADLLKGRVENLERELEIARTNQEHAALEAENSKGEVETLKAKIEGMTQ-----SLRG 2478

Qy   406 LQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDSDCSLVSSTESQLLSDLSATSSR 465
Db  2479 LELDVVTIRSEKENLTNELQKEQERI------------SELEIINSSFENILQE----KEQ 2523

Qy   466 ELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDY 525
Db  2524 EKVQMKEKSSTAMEMLQTQLKELNERVAALHNDQEAC---------KAKEQNLSS 2569

Qy   526 EL----LDTADLPQ---------LESS----LQPVSPGRLDVSESGVLMRRRPARRIL 566
Db  2570 QVECLELEKAQLLQGLDEAKNNYIVLQSSVKGLIQEVEDGKQKLEK----KDEEISRLK 2624

Qy   567 SQVTMLAFQGDALLEQISVIGG---------NLTGIFIHRVTPGSAADQMALRPGT 613
Db  2625 NQIQ--DQEQLVSKLSQVEGEHQLWKEQNLELRNLT-VELEQIQVLQSKNASLQDTL 2679

Qy   614 QIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVA 673
Db  2680 EVLQSSYKNLE---NELELTKMDK-MSFVEKVNKMTAKEFELQREHEMAQKTAELQEE 2734

Qy   674 TSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQ---------GCGCWHAHRVNS 723
Db  2735 LSGE--------KNRLAGELQLLLEEIKSSKDQLKELITLENSELKKSLDCMHKDQVE- 2783

Qy   724 YTMKDTAAHGTIPNYS---------RAQQOLIALIQDMTQQCTV------TRKPSSGGPQKL 770
```

```
Qy   473 SSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDYELLDTAD 532
Db   724 RAADALEQQRCISELKAE------TRSLVEQHKRERKELEEERAGRKGLEARLLQLGE 776
Qy   533 LPQLESSLQPVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTG 592
Db   777 AHQAETE---------------VLRRELAEAMAAQHT----AESECEQL------- 806
Qy   593 IFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCL 652
Db   807 ---------VKEVAAWRDGYE---DSQQEEAQYGAMFQEQ-------------LM 836
Qy   653 SVKVNTDGYKRLLQDLEAKVATSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQG 712
Db   837 TLKEECKARQELQEAKEKVA-GIEHSELQISRQQNKLAEL--HAN-------- 880
Qy   713 CGCWHAHRVNSYTMKDTAAHGTIPNYSRAQQQLIALIQDMTQQCTVTRKPSSGGPQKLV- 771
Db   881 ---LARALQQVQEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRK--AGEQQETAS 934
Qy   772 RIVSMDKAKASPLRLSFDRGQLDPSRME---GSSTCFWAESCLTLVPYTLVWPHRPARPR 828
Db   935 RELVKEPARA----------GDRQPEWLEEQQGRQFC---------------STQA 965
Qy   829 PVLLVPRAVGKILSEKLCLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGGRCWVTR 888
Db   966 ALQAMEREAEOMGNE----LERLRAALMESQGSQOQEERGQQEREVARLTQERG------R 1015
Qy   889 HAVESLMEKNTHALLDVQLDSVCTLHRMDIFPI--VIHVSVNEKMAK-----KLK----- 936
Db  1016 AQADLALEKAARAELEMRLQNALNEQRVEFATIQEALAHALTEKEGKDQELAKLRGLEAA 1075
Qy   937 --KGLQRLGTSEEQLLEAARQEE 957
Db  1076 QIKELEELRQTVKQLKEQLAKKE 1098
```

RESULT   8
US-08-353-700-1
```
;   Sequence 1, Application US/08353700
;   Patent No. 5599919
;   GENERAL INFORMATION:
;   APPLICANT: YEN, TIMOTHY J.
;   APPLICANT: RATTNER, JEROME B.
;   TITLE OF INVENTION: NUCLEIC ACID ENCODING A
;   TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
;   TITLE OF INVENTION: AND METHODS OF USE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
;   STREET: 1601 MARKET STREET, SUITE 720
;   CITY: PHILADELPHIA
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19103-2307
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/353,700
;   FILING DATE: 09-DEC-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: REED, JANET E.
;   REGISTRATION NUMBER: 36,252
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (215) 563-4100
;   TELEFAX: (215) 563-4044
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3248 amino acids
```

```
Db  2784  --KEGKVREEIAEYQLRLHEAEKKHQALLLDTNKQYEVEIQTYREKLTSKEECLSSQKL 2840
Qy   771  VRIVSMDKAKASPLRLSFD-------RGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHR 823
                :|:|     :||:||
Db  2841  ----EIDLLKSSKEELNNSLKATTQILEELKKTKMD----------------NLKYVNQ 2879
Qy   824  PARPRPVLLVPRAVGKI-LSEKLC-LLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSG 881
                 ||  |||:||    |:| |     || :||:|||||||||||||||| ||
Db  2880  LKKEN-----ERAQGKMKLLIKSCKQLEEKEILQKELSQLQAAQEKQKTGTVMDTKVDE 2934
Qy   882  GRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQR 941
           ||:   ::|| | :||       ||| || ::|:       |:||
Db  2935  LTTEI-KELKETLEEKTKEA--DEYLDKYCSL-------LISH--------EKLEKAKEM 2976
Qy   942  LGTSEEQLLEAARQEEGDLDRAPCL 966
           ||     || :| || |::
Db  2977  LETQVAHL--CSQQSKQDSRGSPLL 2999
```

```
RESULT 9
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street  Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match         4.6%;  Score 238;  DB 5;  Length 3248;
Best Local Similarity  19.4%;  Pred. No. 1.3e-10;
Matches  214;  Conservative  192;  Mismatches  393;  Indels  306;  Gaps  48;

Qy    2  GELCRRDSALTALD-------EETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDE 54
            ||     ||:|        :|  |:  ::|||   |   | :|: | :|  | :||
Db  2061  GELDTMSKKTTALDQLSERMKEKT--QELESHQSECLHCI----------------QVAE 2102
```

```
Db  2784  --KEGKVREEIAEYQLRLHEAEKKHQALLLDTNKQYEVEIQTYREKLTSKEECLSSQKL 2840
Qy   771  VRIVSMDKAKASPLRLSFD-------RGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHR 823
Db  2841  ----EIDLLKSSKEELNNSLKATTQILEELKKTKMD----------------NLKYVNQ 2879
Qy   824  PARPRPVLLVPRAVGKI-LSEKLC-LLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSG 881
Db  2880  LKKEN-----ERAQGKMKLLIKSCKQLEEKEILQKELSQLQAAQEKQKTGTVMDTKVDE 2934
Qy   882  GRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQR 941
Db  2935  LTTEI-KELKETLEEKTKEA--DEYLDKYCSL-------LISH--------EKLEKAKEM 2976
Qy   942  LGTSEEQLLEAARQEEGDLDRAPCL 966
Db  2977  LETQVAHL--CSQQSKQDSRGSPLL 2999
```

```
Qy    55  EEVLHSPRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDYTLVTGLQPDVDFSN 114
               |    |       ::||||                       :  :||
Db  2103  AEVKEKTELLQTL--SSDVSLLKD------------------KTH---LQEKLQSLEKD- 2139
Qy   115  FSGLMETSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLG------------- 161
                :|  | |     ::| ||:|||::| ||||   :|:| ||:| |:
Db  2140  ---SQALSLTKC-----ELENQIAQLNKEKELLVKESESLQARLSESDYEKLNVSKAL 2189
Qy   162  -----------LAETRAEGLHQLEADHSRMKREVSA-----HFHEVLR-----------L 194
                       |:   || ||:|||:   |:| |:|     ||:|||
Db  2190  EAALVEKGEFALRLSSTQEE-VHQLRRGIEKLRVRIEADEKKQLHIAEKLKERERENDSL 2248
Qy   195  KDEMLSLSLHYSNALQEKELAASRCRSLQELYLLKQELQRANMVSSCEL-ELQEQSLRT 253
           || ||  |:||:|| ||:|| ::| :||:  ||:|||: |: |::       | ||| |
Db  2249  KDKVENLERELQMSEENQELVLDAENSKAEVETLKTQIE-EMARSLKIFELDLVTLRS 2306
Qy   254  ASDQ--ESGDELNRLKEENEKLRSLTFSLAEKDILE-QSLDEARGSRQELVERIHSLRE 310
              :     :|| |: |:||| |:||  |||| :|| |||:||| :||::|::| :|||
Db  2307  EKENLTKQIQEKQGQLSELDKLLSSFKSLLEEKEQAEIQIKEESKTAVEMLQNQLKELNE 2366
Qy   311  RAVAAERQREQYWEEKEQTL------------------------LQFQKSKMACQLYREK- 346
           |  |||  ||:  |::::|                        |:|:|::|: | ||:
Db  2367  -AVAALCGDQEIMKATEQSLDPPIEEHQLRNSIEKLRARLEADEKKQLCVLQQLKESEH 2425
Qy   347  -VNALQAQVCELQKERDQAYSARDSAQRE1SQSLVEKDSLRRQVFELTDQVCELRTQLRQ 405
            ||| :|||  |:||  ||  : | ||:  ::|  :||:  ||::| :|:|   |  |
Db  2426  HADLLKGRVENLERELEIARTNQEHAALEAENSKGEVETLKAKIEGMTQ------SLRG 2478
Qy   406  LQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSR 465
           |: |    ||:|| :|      |:|:||||:||:  |  |:|   |:     | |:  | |
Db  2479  LELDVVTIRSEKENLTNELQKEQERI--------SELEIINSSFENILQE---KEQ 2523
Qy   466  ELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDY 525
           |:|           ||:||   | |:   |  |||        |       |  |
Db  2524  EKVQMKEKSSTAMEMLQTQLKELNERVAALHNDQEAC-------------KAKEQNLSS 2569
Qy   526  EL----LDTADLPQ---------LESS---LQPVSPGRLDVSESGVLMRRRPARRIL 566
           ||     || |||           || :  | |  |  |::|   ||:::| || |
Db  2570  QVECLELEKAQLLQGLDEAKNNYIVLQSSVKGLIQEVEDGKQKLEK----KDEEISRLK 2624
Qy   567  SQVTMLAFQGDALLEQISVIGG--------------NLTGIFIHRVTPGSAADQMALRPGT 613
           : |      | |: |:| | ||              |:  ||    || :|:|  | |
Db  2625  NQIQ----DQEQLVSKLSQVEGEHQLWKEQNLELRNLT-VELEQKIQVLQSKNASLQDTL 2679
Qy   614  QIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEAKVA 673
           |:  ||  |:|  |: | |     |    | :|     |:   |::|:  || || :||
Db  2680  EVLQSSYKNLE---NELELTKMDK-MSFVEKVNKMTAKETELQREMHEMAQKTAELQEE 2734
Qy   674  TSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQ--------------GCGCWHAHRVNS 723
               :||||:|    | |||::|| | |:| |  | |:              | |: | ||:
Db  2735  LSGE--------KNRLAGELQLLEEIKSSKDQLKELTLENSELKKSLDCMHKDQVE- 2783
Qy   724  YTMKDTAAHGTIPNYS---RAQQQLIALIQDMTQQCTV-----TRKPSSGGPQKL 770
           :| ||  || |||:|    | |   | |  || | |        | |:||  ||:
Db  2784  --KEGKVREEIAEYQLRLHEAEKKHQALLLDTNKQYEVEIQTYREKLTSKEECLSSQKL 2840
Qy   771  VRIVSMDKAKASPLRLSFD-------RGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHR 823
Db  2841  ----EIDLLKSSKEELNNSLKATTQILEELKKTKMD----------------NLKYVNQ 2879
Qy   824  PARPRPVLLVPRAVGKI-LSEKLC-LLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSG 881
Db  2880  LKKEN-----ERAQGKMKLLIKSCKQLEEKEILQKELSQLQAAQEKQKTGTVMDTKVDE 2934
Qy   882  GRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQR 941
Db  2935  LTTEI-KELKETLEEKTKEA--DEYLDKYCSL-------LISH--------EKLEKAKEM 2976
Qy   942  LGTSEEQLLEAARQEEGDLDRAPCL 966
Db  2977  LETQVAHL--CSQQSKQDSRGSPLL 2999
```

```
RESULT 10
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
;   APPLICANT: Zhu, Xueliang
;   APPLICANT: Lee, Wen-Hwa
;   TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Campbell and Flores
;   STREET: 4370 La Jolla Village Drive, Suite 700
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92122
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/328,254
;   FILING DATE: 24-OCT-1994
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/141.239
;   FILING DATE: 22-OCT-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Campbell, Cathryn A.
;   REGISTRATION NUMBER: 31,815
;   REFERENCE/DOCKET NUMBER: P-CJ 1191
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619) 535-9001
;   TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2482 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-328-254-6

Query Match         4.6%;   Score 237;  DB 1;  Length 2482;
Best Local Similarity   19.4%;   Pred. No. 1e-10;
Matches 214;  Conservative 192;  Mismatches 393;  Indels 306;  Gaps 48;

Qy      2 GELCRRDSALTALD------EETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQLDE 54
            ||||                ||| :: : ::|||: :  |
Db   1333 GELDTMSKKTTALDQLSEKMKEKT--QELESHQSECLHCI------------QVAE 1374

Qy     55 EEVLHSPRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPDVYTLVTGLQPDVDFSN 114
            :|| | |: :: ::|||
Db   1375 AEVKEKTELLQTL--SSDVSELLKD----------KTH---LQEKLQSLEKD------ 1411

Qy    115 FSGLMETSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQLQEHLG------- 161
            || ||  : : :: : :|:|||:
Db   1412 ---SQALSLTKC-----ELENQIAQLNKEKELLVKESESLQARLSESDYEKLNVSKAL 1461

Qy    162 --------------LAETRAEGLHQLEADHSRMKREVSA-----HFHEVLR--------L 194
                          | : :::|  ||: :  |
Db   1462 EAALVEKGEFALRLSSTQEE-VHQLRRGIEKLRVRIEADEKKQLHIAEKLKERERENDSL 1520

Qy    195 KDEMLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCEL-ELQEQSLRT 253
            ||:: :| |: :  :: |:: :| ||  :: ::
Db   1521 KDKVENLERELQMSEENQELVILDAENSKAEVETLKTQIE--EMARSLKVFELDLVTLRS 1578

Qy    254 ASDQ--ESGDEELNRLKENEKLRSLTFSLAEKDILE-QSLDEARGSRQELVERIHSLRE 310
            | :||  : |:: |: ||: ::|  :::  ::|||
Db   1579 EKENLTKIQEKQGQLSELDKLLSSFKSLLEEKEQAEIQIKEESKTAVEMLQQNLQLKELNE 1638

Qy    311 RAVAAERQREQYWEEKEQTL------LQFQKSKMACQLYREK---- 346
```

```
Db   1639 -AVAALCGDQEIMKATEQSLDPPIEEHQLRNSIEKLRARLEADEKKQLCVLQQLKESEH 1697
                 ||||  ::       ::     |    :  |     |  :  :: :|
Qy    347 -VNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQ 405
            : :||:  : :: : ||  : : |:|| |:
Db   1698 HADLLKGRVENLERELEIARTNQEHAALEAENSKGEVETLKAKIEGMTQ-------SLRG 1750

Qy    406 LQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSR 465
            : :: :||  :  : :   :::  |
Db   1751 LELDVVTIRSEKENLTNELQKEQERI-------SELEIINSSFENILQE----KEQ 1795

Qy    466 ELVDSFRSSSPAPPSQQSLYKRVAEDFGEEPWSFSSCLEIPEGDPGALPGAKAGDPHLDY 525
            :  |  |    | |  :  |      |
Db   1796 EKVQMKEKSSTAMEMLQTQLKELNERVAALHNDQEAC------KAKEQNLSS 1841

Qy    526 EL----LDTADLPQ---------LESS---LQPVSPGRLDVSESGVLMRRRPARRIL 566
            :|     |:|| |           :::    |  :  :||:|:   ::|::  | : |
Db   1842 QVECLELEKAQLLQGLQGLDEAKNYIVLQSSVNGLIQEVEDGKQKLEK----KDEEISRLK 1896

Qy    567 SQVTMLAFQGDALLEQISVIGG---------NLTGIFIHRVTPGSAADQMALRPGT 613
            : ::| |   | |::::                 :   :   |: ::   :|
Db   1897 NQIQ--DQEQLVSKLSQVEGEHQLWKEQNLELRNLT-VELEQKIQVLQSKNASLQDTL 1951

Qy    614 QIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNTDGYKRLLQDLEEAKVA 673
            |: ::    :   |: : |  :|::|
Db   1952 EVLQSSYKNLE---NELELTKMDK-MSFVEKVNKMTAKETELQREMHEMAQKTAELQEE 2006

Qy    674 TSGDSFYIRVNLAMEGRAKGELQVHCNEVLHVTDTMFQ----------GCGCWHAHRVNS 723
            || |   |           |  : |
Db   2007 LSGE----KNRLAGELQLLEEIKSSKDQLKELTLENSELKKSLDCMHKDQVE------- 2055

Qy    724 YTMKDTAAHGTIPNYS----RAQQQLIALIQDMTQQCTV----TRKPSSGGPQKL 770
            :::|  :                     ::|  |:||  :|           : :|
Db   2056 ---KEGKVREEIAEYQLRLHEAEKKHQALLDTNKQYEVEIQTYREKLTSKEECLSSQKL 2112

Qy    771 VRIVSMDKAKASPLRLSFD-----RGQLDPSRMEGSSTCFWAESCLTLVPYTLVWPHR 823
            |    ::  | :  | : :       : || |   : :     ::| :   |  : |
Db   2113 ----EIDLLKSSKEELNNSLKATTQLEELKKTKMD--------NLKYVNQ 2151

Qy    824 PARPRPVLLVPRAVGKI-LSEKLC-LLQGFKKCLAEYLSQEYEAWSQRGDIIQEGEVSG 881
            | :| :|| |  | |:   | ||      ||   | |   |   |   |    |: | |
Db   2152 LKKEN-----ERAQGKMKLLIKSCKQLEEEKEILQKELSQLQAAQEKQKTGTVMDTKVDE 2206

Qy    882 GRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVIHVSVNEKMAKKLKKGLQR 941
            |  |  : |  |: :    |:|  | || |   :   :|| : :   |:| |:| |  :|
Db   2207 LTTEI-KELKETLEEKTKEA--DEYLDKYCSL-----LISH------EKLEKAKEM 2248

Qy    942 LGTSEEQLLEAARQEEGDLDRAPCL 966
            |    ::  :  | :: | : |
Db   2249 LETQVAHL--CSQQSKQDSRGSPLL 2271
```

```
RESULT 11
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
;   APPLICANT: Kaibuchi, Kozo
;   APPLICANT: Iwamatsu, Akihiro
;   APPLICANT: Nakano, Takeshi
;   APPLICANT: Ito, Masaaki
;   APPLICANT: Takahashi, No. 5906819uaki
;   TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 3000 K Street, N.W., Suite 500
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20007-5109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
```

```
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/685,576
      FILING DATE: 24-JUL-1996
      CLASSIFICATION: 435
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 7-325129
      FILING DATE: 20-NOV-1995
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 8-17150
      FILING DATE: 05-JAN-1996
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 8-131206
      FILING DATE: 26-APR-1996
   ATTORNEY/AGENT INFORMATION:
      NAME: Bent, Stephen A.
      REGISTRATION NUMBER: 29,768
      REFERENCE/DOCKET NUMBER: 16887/843
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202)672-5300
      TELEFAX: (202)672-5399
      TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 1388 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
   MOLECULE TYPE: protein
US-08-685-576-1

Query Match           4.2%;    Score 216.5;   DB 2;   Length 1388;
Best Local Similarity 19.6%;   Pred. No. 2.3e-09;
Matches 151;   Conservative 129;   Mismatches 292;   Indels 199;   Gaps 28;

Qy     6 RRDSALTALDEE-TLWEMMESHRHRIVR--CICPSRLTPYLRQAKVLCQLDEEEVLHSPR 62
         :  :   ||:    :|| :   :       |   :|:|  :  :|  |: :|      :
Db   476 RLEKVAKELEEITLRKNVESTLRQLEREKALLQHKNAEYQRKA------DHEADKKRN 528

Qy    63 LTNSAMRAGHLLDLLKTRGKNGAIAF-------------------------------- 88
         |    |  || :| :  || :|: :|
Db   529 LENDVNSLKDQLEDLKKRNQNSQISTEKVNQLQRQLDETNALLRTESDTAARLRKTQAES 588

Qy    89 ---LESLKFHNPDVY----TLVTG-LQPDVDFSNFSGLMET------------SKLTECLAGAI 132
            : |:||  : ::        | |  :| | |   |:|:            |  |  | | |
Db   589 SKQIQQLESNNRDLQDKNCLLETAKLKLEKEFINLQSVLESERRDRTHGSEIINDLQGRI 648

Qy   133 GSLQEELNQEKGQKEVLLRRCQ----QLQEHLGLAETRAEGL-----HQLEADHSRMKRE 183
         ||| : :| ||  :|||| |        | ||  |  | | |     | || : |:|  :
Db   649 SGLEEDVKNGK----ILLAKVELEKRQLQERFTDLEKEKNNMEIDMTYQLKVIQQSLEQE 704

Qy   184 VSAHFHEVLRLKDE--MLSLSLHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVS 240
          : |  |   | ::   :|   |  : ::|  ||:  :|  || :|  || || |:
Db   705 ETEHKATKARLADKNKIYESIEEAKSEAMKEMEKKLSEERTLKQKVENLLLEAEKRCSIL 764

Qy   241 SCELELQEQSLR---------- 265
         :  |  |:|: |
Db   765 DCDLKQSQQKINELLKQKDVLNEDVRNLTLKIEQETQKRCLJTQNDLKMQTQQVNTLKMSE 824

Qy   266 -RLKEENEKLRSLTFSLAEKDI-LEQSLDEARGSRQELVERIHSLRERAVAAEROREQYW 323
          || : || :|  | ||  |   |:|:  |   |  :|  ::| :: :|    :  :
Db   825 KQLKQENNHLLEMKMSLEKQNAELRKERQADAGQMKELQDQL---------------EAEQYF 872

Qy   324 EEKEQTLLQFQKSKMACQLYREKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKD 383
          ||  | | |::|     | :|  :|| | |  | |:     :|  |::|::
Db   873 STLYKT--QVRELKEECEEKTKLCKELQQKKQELQDERDSL----AAQLEITLTKADSE 925

Qy   384 SLRRQVFELTDQVCELRTQLRQLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDDS 443
          :|| |  |   :|| |  ::|  |  || |
Db   926 QLARSIAE--------EQYSDLEKEK--IMKELEIKEMMARHKQELT-------EKDA 966
```

```
Qy   444 DCSLVSSTESQLLSDLS--ATSSRELVDSFRSSSPAPPSQQSLYKRVAEDFGEPWSFSS 501
          |        |   :|:   | |  :| | |   | |   :| |:|     |   | |
Db   967 TIASLEETNRTLTSDVANLANEKEELNNKLKE------AQEQLSRLKDEI-------- 1011

Qy   502 CLEIPEGDPGALPGAKAGDPHLDYELD----------TADLPQLESSLQPVSPG----R 547
          :    | |  | |  |  |                 :|   |:: | |
Db  1012 --------SAAAIKAQFEKQLLTERTLKTQAVNKLAEIMNRKEPVKRGNDTDVRR 1058

Qy   548 LDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFIHRVTPGSAADQM 607
          |    | | |:|  | |  |:      | :| ||  |   : :|  : |    |
Db  1059 KEKENRKLHMELKSEREKLTQ-QMIKYQKELNEMQAQIAEESQIRIELQMTLDSKDSDIE 1117

Qy   608 ALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVNT 658
          :|  |  |:  |   |  |   |   ::|     |  | | |    | | |
Db  1118 QLRSQLAHIGLDSSS--IGSGPDTEADDGFP-ESRLEGWLSLPVRNNT 1165
```

```
RESULT 12
US-09-104-324B-4
; Sequence 4, Application US/09104324B
; Patent No. 6232460
; GENERAL INFORMATION:
;    APPLICANT: T reci, Ozlem; Sahin, Ugur; Pfreundschuh, Michael
;    TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
;    TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample
;    TITLE OF INVENTION: No. 6232460mal Cells
;    NUMBER OF SEQUENCES: 4
;    CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Fulbright & Jaworski LLP
;    STRET: 666 Fifth Avenue
;    CITY: New York City
;    STATE: New York
;    ZIP: 10103
;    COMPUTER READABLE FORM:
;    MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
;    COMPUTER: IBM
;    OPERATING SYSTEM: PC-DOS
;    SOFTWARE: Wordperfect
;    CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/09/104,324B
;    FILING DATE: 25-June-1998
;    CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: 08/892,702
;    FILING DATE: 15-July-1997
;    ATTORNEY/AGENT INFORMATION:
;    NAME: Hanson, No. 6232460man D.
;    REGISTRATION NUMBER: 30,946
;    REFERENCE/DOCKET NUMBER: LUD 5491
;    TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (212) 318-3000
;    TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 4:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 976 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
US-09-104-324B-4

Query Match           4.1%;    Score 210.5;   DB 4;   Length 976;
Best Local Similarity 22.8%;   Pred. No. 4.2e-09;
Matches 119;   Conservative 76;   Mismatches 189;   Indels 139;   Gaps 20;

Qy    12 TALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKV-----------------------L 49
          | || | |:|:|       |    |: :  | ::
Db   303 TKLQSENLKQSIEKQHH------------------------LTKELEDIKVSLQRSVSTQKALEEDLQIATKTI 352

Qy    50 COLDEEE------------------VLHSPRLTNSAMRAGHLLDLLKTRGKNGAIAFLESLKFHNPD 98
          | |:||                   :       :
Db   353 CQLTEEKTQMEESNKARAAHSFVVTEFETTVCSLEELLRTE-------QQRLEKNEDQ 404

Qy    99 VYTLIVTGLQPDVDFSNFSGLMETSKLTECLAGAIGSLQEELNQEKGQKEVLLRRCQQL-- 156
```

Best Local Similarity 20.8%; Pred. No. 9.3e-09;
Matches 105; Conservative 105; Mismatches 184; Indels 110; Gaps 20;

```
Qy     40 TPYLRQAKVLCQLDEEVLHSP--RLTNSAMRAGHLLDLLKTR--GKNGAIAFLE---  90
             : :||:  : ||  .||. |||     ||| :|| |:::| :||   :|
Db    361 TKALSLARAL----EEALEAKEELERTNKMLKA-EMEDLVSSKDDVGKN--VHELEKSK 412

Qy     91 -SLKFHNPDVTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGSLQEELNQEKGQKEVL 149
            :||:: |: : | ||:| .|| : :||: |  ::::|| :||:||||: ||||:
Db    413 RALETQMEEMKTQLEELEDELQASEDAKL---RLEVNMQALKGQFERDLQARDEQNE-- 466

Qy    150 LRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRLKDEMLSLSHYSNAL 209
            |::|||  : ||   ::|||    |||  ::|| :| :| ::  :  :|: :|
Db    467 -EKRRQLQRQLHEYET---------ELEDERNERALAAAKK--KLEGDLKDELQADSAI 515

Qy    210 QEKELAASRCRSLQEELYLLKQELQ-----------RANMVSSCELELQEQSL       251
            ::|:|: || |:|  |::  :||            : :|  ::|||||:||
Db    516 KGREEAIKQLRKLQAOMKDFQRELEDARASRDEIFATAKENEKKAKSLEADLMQLQEDLA 575

Qy    252 RTASDQESGDELNRLKEE-----------NEKLRSLTFSLAEKDILEQSLDEARGSRQE 300
             ||::::| :|:|:||               :|| | |:|| | |::: |:|: :|
Db    576 AAERARRKQADLEKEELAEELASSLSGRNALQDEKRRLEARIAQ---LEEELEEEQGNMEA 632

Qy    301 LVERIHSLRERA------VAAERQREQYWEEKEQTLLQFQKSKMACQL--------YR  344
             |  ||:|:||       | | :|| ||||:| |||: |:|.|  :|        |
Db    633 MSDRVRKATQQAEQLSNELATERSTAQKNESARQQ-LERQNKELRSKLHEMEGAVKSKFK 691

Qy    345 EKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLR 404
             ::  || |   |::|: :| |  |: :|| :     || : :| ||:::|   :|:|
Db    692 STIAALEAKIAQLEEQVEQEAREKQAA---TKSLKQKDKKLKEILLQVEDERKMAEQYK 747

Qy    405 QLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDL-SATS 463
            ::|:: : :|| | ||  | |:||:::      : |: ::| || ::|:|  |  :|
Db    748 EQAEKGNARVKQLKRQLEEAEESQR---------------INANRRKLQRELEDATE 790

Qy    464 SREL---VDSFRSSSPAPPSQQS 483
            |:||      :  |  | |:::
Db    791 SNEAMGREVNALKSKLRGPPPQET 814
```

RESULT 14
US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
;   APPLICANT: Liu, Pu
;   APPLICANT: Collins, Francis S.
;   APPLICANT: Siciliano, Michael J.
;   APPLICANT: Claxton, David
;   TITLE OF INVENTION: Markers for Detection of Chromosome 16
;   TITLE OF INVENTION: Rearrangements
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
;   STREET: P.O. Box 828
;   CITY: Bloomfield Hills
;   STATE: MI
;   COUNTRY: USA
;   ZIP: 48303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/742,923A
;   FILING DATE: No. 5869611ember 1, 1996
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Smith, DeAnn F.
;   REGISTRATION NUMBER: 36683
;   REFERENCE/DOCKET NUMBER: 2115-00869DVC
;   TELECOMMUNICATION INFORMATION:

```
Db    405 LKIITMELQ----KKSSELEEMTKLTNNKEVEL----EELKKVLGEKETLLYENKQFEK 455
            : : |::|||     | : | :| :|:| :|:||         :|  ::::| |||:|
Qy    157 ------QEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRLKDEMLSLSHYSN     207
                  :|:|| | |:| | |::| ||| : : |   |  |   | |:|| : :
Db    456 IAEELKGTEQELIGLLQAREKEVIDLEIQLTAITTSEQYYSKEVKDLKTEL------ENE 509
```

```
Qy    208 ALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELELQEQSLRTASDQESGDEELNRL 267
            ||:||   : | :|:|| |: :|||      ||: :|| ||:|     :| | :|||
Db    510 KLKNTEL-TSHCNKLSLENKELTQE------TSDMTLELKNQ---------QEDINNN 551
```

```
Qy    268 KEENEKLRSLTFSLAE---------KDILEQSLDEARGSRQELVERIHSLRERAVAA 315
            |:||:||:|| :|:||         ::|| |:|::| | :: |  |  ||:: ::|
Db    552 KKQDERMLKQIENLQETETQLRNELEYVREELKQKRDEVKCKLDKSEENCNNLRKQV--- 608
```

```
Qy    316 ERQREQYWEEKEQTLLQFQKSKMA----CQLYREKVNALQAQVCELQKERDQAYSARDSA 371
            | ||| ||| |::||:   ::|:     | || :|: |   | | :||
Db    609 -ENNKYIEELQQENKALLKKGTAESKQLNVYEIKVNKLE--LELESAKQFGEITDTY 664
```

```
Qy    372 QREISQSLVEKDSLRRQVFE---LTDQVCELRTQL-RQLQ--AEPPGVLKQEARTREPC 424
            |:|| :||:| |: | |||    | ||    | ||  | |  || :| |:|   :| |
Db    665 QKEIEDKKISEENLLEEVEKAKVIADEAVKLQKEIDKRCQHKIAEMVALMEKHKHQYDKI 724
```

```
Qy    425 PREKQRLVRMHAICPRDDSDCSLVSSTESQLLSDLSATSSREL 467
            | :||  : |  |   :  |  |:|||:| |:|| ::
Db    725 IEER-------------------DSELGLYKSKE-QEQSSLRASLEIEL 753
```

RESULT 13
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
;   APPLICANT: Liu, Pu
;   APPLICANT: Collins, Francis S.
;   APPLICANT: Siciliano, Michael J.
;   APPLICANT: Claxton, David
;   TITLE OF INVENTION: Markers for Detection of Chromosome 16
;   TITLE OF INVENTION: Rearrangements
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
;   STREET: P.O. Box 828
;   CITY: Bloomfield Hills
;   STATE: MI
;   COUNTRY: USA
;   ZIP: 48303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/533,306A
;   FILING DATE: September 25, 1995
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Smith, DeAnn F.
;   REGISTRATION NUMBER: 36683
;   REFERENCE/DOCKET NUMBER: 2115-00869COB
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (810) 641-1600
;   TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 816 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-533-306A-6

Query Match   4.0%; Score 205; DB 2; Length 816;

```
;      TELEPHONE: (810) 641-1600
;      TELEFAX:   (810) 641-0270
;  INFORMATION FOR SEQ ID NO: 6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 816 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-742-923A-6

Query Match      4.0%;    Score 205; DB 2;  Length 816;
Best Local Similarity  20.8%;  Pred. No. 9.3e-09;
Matches  105;  Conservative  105;  Mismatches  184;  Indels 110;  Gaps  20;

Qy    40  TPYLRQAKVLCQLDEEEVLHSP---RLTNSAMRAGHLLDLLKTR---GKNGAIAFLE---   90
              :  :  |||  :  |   :::  |||   :   ||
Db   361  TKALSLARAL----EEALEAKEELERTNKMLKA-EMEDLVSSKDDVGKN--VHELEKSK  412

Qy    91  -SLKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGSLQEELNQEKGQKEVL  149
             : |  :|  :  |   |::|:::  |  :  ||   :  |   :| |  :  :: |  ||
Db   413  RALETQMEEMKTQLEELEDELQASEDAKL----RLEVNMQALKGQFERDLQARDEQNE--  466

Qy   150  LRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRLKDEMLSLSLHYSNAL  209
             :  :  ||||   ||   | | | :   ::| :  |   |  :    :| |  ::
Db   467  -EKRRQLQRQLHEYET------------ELEDERNERALAAAAKK--KLEGDLKDLELQADSAI  515

Qy   210  QEKELAASRCRSLQEELYLLKQELQ----------RANMVSSCELELQEQSL          251
             | :  ||:::|::::::||  :::||          |  |::   || | ||
Db   516  KGREEAIKQLRKLQAQMKDFQRELEDARASRDEIFATAKENEKKAKSLEADLMQLQEDLA  575

Qy   252  RTASDQESGDEELNRLKEE-----------NEKLRSLTFSLAEKDILEQSLDEARGSRQE  300
             | | |||  ||   |        :  ||   :  |:   :  |
Db   576  AAERARKQADLEKEELAEELASSLSGRNALQDEKRRLEARIAQ---LEEELEEEQGNMEA  632

Qy   301  LVERIHSLRERA-------VAAERQREQYWEEKEQTLLQFQKSKMACQL----------YR  344
             :  |  ::|   :||     ||                        :|  ||
Db   633  MSDRVRKATQQAEQLSNELATERSTAQKNESARQQ-LERQNKELRSKLHEMEGAVKSKFK  691

Qy   345  EKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLR  404
             :  | ||:  :| ||:  :|   :  |   :||
Db   692  STIAALEAKIAQLEEQVEQEAREKQAA----TKSLKQKDKKLKEILLQVEDERKMAEQYK  747

Qy   405  QLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDSDCSLVSSTESQLLSDL-SATS  463
             ||  |    |     |    |  ||| ||:   :  |   :  |
Db   748  EQAEKGNARVKQLKRQLEEAEEESQR-----------------INANRRKLQRELDEATE  790

Qy   464  SREL----VDSFRSSSPAPPSQQS  483
             |   |    |  ||:   |   :  ::
Db   791  SNEAMGREVNALKSKLRGPPPQET  814

RESULT 15
US-08-533-306A-4
; Sequence 4, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
;    APPLICANT: Liu, Pu
;    APPLICANT: Collins, Francis S.
;    APPLICANT: Siciliano, Michael J.
;    APPLICANT: Claxton, David
;    TITLE OF INVENTION: Markers for Detection of Chromosome 16
;    TITLE OF INVENTION: Rearrangements
;    NUMBER OF SEQUENCES: 14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
;      STREET: P.O. Box 828
;      CITY: Bloomfield Hills
;      STATE: MI
;      COUNTRY: USA
;      ZIP: 48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
```

```
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/533,306A
;      FILING DATE: September 25, 1995
;      CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Smith, DeAnn F.
;      REGISTRATION NUMBER: 36683
;      REFERENCE/DOCKET NUMBER: 2115-00869COB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (810) 641-1600
;      TELEFAX:   (810) 641-0270
;  INFORMATION FOR SEQ ID NO: 4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 885 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-533-306A-4

Query Match      4.0%;    Score 205; DB 2;  Length 885;
Best Local Similarity  20.8%;  Pred. No. 1.1e-08;
Matches  105;  Conservative  105;  Mismatches  184;  Indels 110;  Gaps  20;

Qy    40  TPYLRQAKVLCQLDEEEVLHSP--RLTNSAMRAGHLLDLLKTR---GKNGAIAFLE---   90
Db   430  TKALSLARAL----EEALEAKEELERTNKMLKA-EMEDLVSSKDDVGKN--VHELEKSK  481

Qy    91  -SLKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTECLAGAIGSLQEELNQEKGQKEVL  149
Db   482  RALETQMEEMKTQLEELEDELQASEDAKL---RLEVNMQALKGQFERDLQARDEQNE--  535

Qy   150  LRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVSAHFHEVLRLKDEMLSLSLHYSNAL  209
Db   536  -EKRRQLQRQLHEYET------------ELEDERNERALAAAAKK--KLEGDLKDLELQADSAI  584

Qy   210  QEKELAASRCRSLQEELYLLKQELQ----------RANMVSSCELELQEQSL          251
Db   585  KGREEAIKQLRKLQAQMKDFQRELEDARASRDEIFATAKENEKKAKSLEADLMQLQEDLA  644

Qy   252  RTASDQESGDEELNRLKEE-----------NEKLRSLTFSLAEKDILEQSLDEARGSRQE  300
Db   645  AAERARKQADLEKEELAEELASSLSGRNALQDEKRRLEARIAQ---LEEELEEEQGNMEA  701

Qy   301  LVERIHSLRERA-------VAAERQREQYWEEKEQTLLQFQKSKMACQL----------YR  344
Db   702  MSDRVRKATQQAEQLSNELATERSTAQKNESARQQ-LERQNKELRSKLHEMEGAVKSKFK  760

Qy   345  EKVNALQAQVCELQKERDQAYSARDSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLR  404
Db   761  STIAALEAKIAQLEEQVEQEAREKQAA----TKSLKQKDKKLKEILLQVEDERKMAEQYK  816

Qy   405  QLQAEPPGVLKQEARTREPCPREKQRLVRMHAICPRDSDCSLVSSTESQLLSDL-SATS  463
Db   817  EQAEKGNARVKQLKRQLEEAEEESQR-----------------INANRRKLQRELDEATE  859

Qy   464  SREL----VDSFRSSSPAPPSQQS  483
Db   860  SNEAMGREVNALKSKLRGPPPQET  883
```

Search completed: June 13, 2002, 09:24:20
Job time: 153 sec